THE UNIVERSITY
of EDINBURGH

UNDERSTANDING GENOMIC PREDICTION IN CHICKENS

JOANNA JADWIGA ILSKA

Thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

2014

Declaration

I hereby declare that the presented work is my own work and any assistance has been acknowledged. The work submitted has not been used for any other degree or professional qualification.

Joanna Ilska

22 September 2014

ACKNOWLEDGMENTS

I would like to dedicate this thesis to my parents. Although they had hoped I would continue in their footsteps and become a scholar in humanistic sciences, they have accepted my revolutionary (for our family) decision to branch out into the world of animal science and have supported me on every step of my studies. From early childhood, my parents taught me how to marvel at the world surrounding us, and how to appreciate the work of others, particularly in written form. Throughout my studies, although separated by several thousand miles, they were with me in their thoughts at all times. Knowing this gave me both comfort, as I knew they cheered for me, and motivation, as they always believed I could be "the best". I appreciate your support very much, Mum and Dad, and I am very happy to dedicate this work to you - it would not have happened without you!

I also thank my siblings and their spouses. Kuba and Renata, you were there when I was struggling with student's life. Ewa and Iacovos, you have gone through the PhD process not so long ago, and your feedback from the viewpoint of a student was greatly appreciated! Marysia and Jawor, your example of hard work in extreme conditions has set a very good example and made me think that I'm still having an easy life! To all my nieces and nephews, you have been and always will be a spark of joy, bringing a smile to my face no matter what.

Special thanks needs to go to my fiancé Chris. Through ups and downs, you have been my anchor, bringing me back to the safe shore. Your presence calms me down and your faith in me gives me wings. Only with your tolerance and patience I can afford to both work and indulge in my passion for animals! I love you very much!

When I first started my PhD I was feeling like most fresh students, a bit lost and lacking confidence. The team at the Roslin Institute has quickly made me feel very welcome. I am indebted to Riccardo Pong Wong, for your help with computational "stuff", Valentina Riggio, for words of encouragement, Serap Gonen, Smaragda Tsaridou, John Stainton, Amy Rowlatt, Fiona Milton, Andrea Wilson and many more, for advice, support and good laughs. I am also very grateful for the guidance and extremely useful feedback I received from my thesis committee: Dr Ian Dunn, Prof Dave Burt and Prof Steve Bishop. I would also like to thank my PhD Thesis Defence Examiners: Prof Paul Hocking and Dr Chris Sorensen. Thank you for making this one last step of my studies an enjoyable and confidence instilling experience.

The fraternity of scientists extends beyond our University, or even country. The staff and students at the Norwegian University of Life Sciences have made me feel very welcome during my visit to Ås. This visit gave me the unique opportunity to work with Prof Theo Meuwissen, who very kindly shared his knowledge with me.

Special thanks need to go to Isobel Stewart and Alan Grey, who greatly helped in the development of software capable of dealing with my data. Thanks to them I was able to run analyses on the largest dataset in our group!

My own computational skills have improved by a thousand-fold thanks to my industrial supervisor, Dr Andreas Kranis. He has supported me in the very first steps in programing. As I am not a natural in logical thinking, programing seemed like magic to me for a very long time. Now, I conjure scripts and programs which are far more advanced than what I ever expected. So I would like to thank Andreas for not only showing me how to make life of a scientist easier by using programing skills, but also for showing me that my previously perceived personal limitations could be overcome.

However, my deepest and most sincere gratitude must go to Prof John Woolliams. After realizing that work in the commercial sector was not for me, he was the first person I turned to for advice. Prof Woolliams is known to all animal geneticists in the world. His knowledge and intelligence has been one of the forces driving the animal genetic improvement across countries within last decades. Just that would be enough for any aspiring geneticist to want to be his student. But only those that have actually worked with John can fully appreciate how great of a teacher he is. There is no doubt in my mind that I could not have gone through my project without John's guidance. He was there to push me when I would get stuck. He was there to support me in an unobtrusive manner at public functions. He would always provide me with kind and fair criticism. Although our meeting in the early years were sometimes painful, with my knowledge magically disappearing whenever John would ask a question, it was these meetings that would give me thorough understanding of the problems at hand. Despite sometimes silly mistakes I was making, John was always patient and supportive. I am also very grateful to John for spending some time with me talking about dog genetics, helping me develop my passion with support of his extensive knowledge and experience.

The past four years were an amazing adventure. I have enjoyed being a student for 9 long years. Now, having finished my PhD I finally feel ready to move on into the "adult" life.

Table of Contents

Page

List of Tables Page

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

APPENDIX 1

List of Figures Page

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

ABSTRACT

Genomic prediction (GP) is a novel tool used for prediction of EBVs by using molecular markers. Within the last decade, GP has been widely introduced into routine evaluations of cattle, pig and sheep populations, however, its application in poultry has been somewhat delayed, and studies published to date have been limited in terms of population size and marker densities. This study shows a thorough evaluation of the benefits that GP could bring into routine evaluations of broiler chickens, with particular attention given to the accuracy and bias of Genomic BLUP (GBLUP) predictions. The data used for these evaluations exceeds the numbers of both individuals and marker genotypes of previously published reports, with the studied population consisting of up to 23,500 individuals, genotyped for up to 600K SNPs.

The evaluation of GBLUP is preceded by evaluation of the variance components using traditional restricted maximum likelihood (REML) approach sourcing information from phenotypic records and pedigree, which provide an up to date reference for the estimates of variance components. Chapter 2 tested several models exploring potential sources of genetic variation and revealed the presence of significant maternal genetic and environmental effects affecting several commercial traits. In Chapter 3, a vast dataset containing 1.3M birds spread over 24 generations was used to evaluate changes in genetic variance of juvenile body weight and hen housed production over time. The results showed a slow but steady decline of the variance. Chapter 4 provided initial estimates of the accuracy and bias of genomic predictions for several sex-limited and fitness traits, obtained for a moderately sized population of over 5K birds, genotyped with 600K Affymetrix Axiom panel from which several chips of varying marker densities were extracted. The accuracy of those predictions showed a great potential for most traits, with GBLUP performance exceeding that of traditional BLUP. Chapter 5 investigated the effect of marker choice, with two chips used: one created from GWAS hits and second from evenly spaced markers, both with constant density of 27K SNPs. The two chips were used to calculate genomic relationship matrices using Linkage Analysis and Linkage Disequilibrium approaches. Markers selected through GWAS performed better in Linkage Analysis than in Linkage Disequilibrium approach. The optimum results however were found for relationship matrices which regressed the genomic relationships back to expected pedigree-based relationships, with the best regression coefficient dependent on the chip used. Chapter 6 formed a comprehensive evaluation of the

utility of GBLUP in a large broiler population, exceeding 23,500 birds genotyped using 600K Affymetrix Axiom panel. By splitting the data into variable scenarios of training and testing populations, with several lower density chips extracted from the full range of genotypes available, the effect of population size and marker density was evaluated. While the latter proved to have little effect once 20K SNPs threshold was exceeded, the effect of the population size was found to be the major limiting factor for the accuracy of EBV predictions. The discrepancy between empirical results found and theoretical expectations of accuracy based on the similar genomic and population parameters showed an underestimation of the previously proposed requirements.

Chapter 1

GENERAL INTRODUCTION

GENETIC IMPROVEMENT IN BROILER PRODUCTION

1. POULTRY PRODUCTION

One of the most important issues faced by the modern world revolves around ensuring food security for the global population. The world population has recently exceeded 7 billion people, more than twice the number from 1960's. It is expected that this growth will continue, with the population numbers likely to reach 10 billion by 2050 (FAO, 2014). This global explosion unavoidably poses challenges to food producers, both in terms of quantity and quality of their products. Current estimates of global food consumption indicate that animal products provide nearly 40% of daily protein intake per capita (FAO, 2014). Among the livestock species, poultry constitutes the largest sector, with over 99 million tonnes of poultry meat produced in 2011 (FAO, 2014). From a perspective of a livestock producer, poultry possesses several benefits over other species, for example they have short generation interval, fast growth rate, multiple offspring per dam and due to their small size can be easily accommodated in limited spaces. These factors result in a relatively low price for poultry products, which in turn makes them an attractive option for consumers. Thus, the popularity of poultry production has been consistently increasing over the last century, with the percentage share of poultry in the total world meat production increasing from just over 12% in 1961 to 34% in 2011, as shown in Figure 1.1. The popularity of the poultry is particularly apparent in developing countries, e.g. in Asia, where poultry constitutes as much as 85% of all livestock species (FAO, 2014).

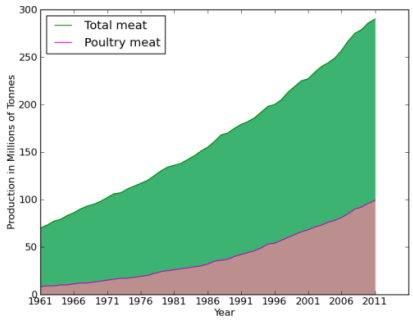

Figure 1.1 World poultry meat production as a part of total meat production.

2. PROGRESS USING BLUP

The huge demand for poultry products has been met by breeders primarily through an increase in the numbers of chickens produced, from $3.9x10^6$ in 1961, to $21.9x10^6$ in 2012 (FAO, 2014). However, the increase in numbers has been also accompanied by improvements in the meat production per bird. It is estimated that the broiler weight gain increases at a rate of 84g/year thanks to the genetic improvement of the lines and developments in management routines (Havenstein et al., 2003). Although the latter is unquestionably an important factor of the commercial production, the genetic advancement of broilers has been proven in numerous studies accounting for changes in the birds' environment. For example, a comparison between modern broilers and an experimental line originating from Athens-Canadian Randombred Control line established in 1957, with both lines fed on a diet typical for 1950's poultry diets, showed that modern birds are over three times larger than the unselected counterpart at the same age (Havenstein et al., 2003). In another experiment, a divergent selection for body weight at the age of 8 weeks carried out over 54 generations showed the spectacular success in both increasing and decreasing the weight. Starting from 800g which was an average body weight recorded in the founding stock of this experiment, the selection resulted in over 1,700g of body weight in birds selected for increased weight and as little as 150g in birds selected for low weight (Dunnington et al., 2013).

Since the domestication times, genetic improvement of livestock has been achieved by selecting the individuals that exhibited the most desirable characteristics. Initially the selection was based on the observable characteristics of the selection candidates, however this method had limited success in traits with low heritability, or in traits where only one sex exhibited the phenotype, e.g. egg production. Due to their individual-centred nature, these

early evaluation methods did not specifically utilize the familial information of pedigree breeders, where the records of relatives could provide insight on the missing record of the selection candidate. The progress in breeding programs has been thus greatly facilitated when Henderson (1950) developed a method called Best Linear Unbiased Prediction (BLUP), which exploited all available sources of information, including the information from the selection candidate's relatives, with appropriate weights given to the records of the relatives, depending on their relationship with the candidate and depending on the heritability (h^2) of the trait in question. The properties of BLUP are neatly listed in its name: Best - signifies the maximized correlation between estimated and true breeding values (TBVs), Linear - explains that the predictors are linear functions of the observed phenotypes, Unbiased - means that the expected deviation of the estimated value (e.g. EBV) from the true value (true breeding value, TBV) is zero, and Prediction - indicates that the end product yields the predictors of the random effects, i.e. breeding values (Mrode, 2005).

The BLUP methodology refers the phenotypic records to the covariances between relatives, quantified through the numerator relationship matrix (**A**), thus allowing prediction of Estimated Breeding Values (EBVs) even for individuals without phenotypes. Concurrently with the use of family links, BLUP introduced a significant improvement to the predictions of EBVs by accounting for the environmental effects, possible through the use of mixed linear models (MLM) on which this method was constructed.

The relationship matrix **A** used in BLUP is symmetric and non-singular, and represents the expected covariance arising from the allele sharing between relatives, based on statistical expectations of allele segregation from a base population. The methodology for calculating the **A** matrix was first presented by Wright (1922) and its coefficients can be quantified as twice the coefficients of coancestry. The **A** matrix contains the relationships between individuals recorded in formalized pedigree structure, thus it has a finite size. The founders of the pedigree are treated as a base population, and are assumed to be unrelated, non-inbred and are assumed to originate from random mating. Thus, the part of the **A** matrix relating to the base population is an identity matrix, where the diagonals consist of 1s and represent the relationship of the individual with itself, while off-diagonals are 0s. For other individuals in the pedigree, excess over the 1s in diagonal coefficients quantifies the inbreeding coefficient.

Albeit that BLUP methodology has allowed spectacular improvement in chicken production, it has some limitations. It has been termed as a "black box" approach, where predictions on the genetic value of an animal are constructed without the evaluation of the actual genes and alleles it carries. This pertains particularly to the assumed relationship coefficients. For

example, in absence of inbreeding, full siblings are expected to share half of the alleles inherited from the parents, half siblings are expected to share a quarter etc. In reality this covariance varies between particular pairs of individuals, with exception of parent offspring pairs, where the descendant will always inherit a half of its genetic constitution from the parent (VanRaden, 2007). The distribution of the proportion of alleles that are indeed identical between two individuals depends on the number of loci that are considered. The expected values are based on the infinitesimal model, where the number of loci affecting a trait is, as name suggests, infinite. However, as the number of loci considered decreases, the standard devation of the allele sharing increases, e.g. standard deviation of the allele sharing between two individuals at just one locus is estimated at 35.4% (VanRaden, 2007). The departure from the expected value arises as a result of the random sampling of the paternal and maternal chromosomes in the gamete formation, with added complexity of possible recombinations. The Mendelian sampling contributes to approximately half of the genetic variance among the selection candidates, with the other half determined by the genetic variance of the selection candidate's parents (van der Werf, 2006). This Mendelian sampling variation is renewed in every generation and, under the infinitesimal model, is not affected by selection. The expected impact of Mendelian sampling defines the genetic contributions of particular individuals to the population (Woolliams et al., 1999). Thus, analyses based directly on the genotypes of individuals pose an attractive development to the genetic improvement programs. While the estimates of genetic variance obtained using traditional and genomic methods should be similar, as it is a parameter of a population, therefore including large number of individuals and genes (Visscher et al., 2006), direct information on the genotype of an individual has the potential to bring a significant increase in accuracy of EBV prediction (Hayes et al., 2009b).

Estimation of the animal's breeding value based on the genotype of this individual, rather than on the phenotypic measurement, also presents other benefits. Aside from an expected increased accuracy of prediction which leads to improved response to selection, such calculations can also lead to a shortening of the generation interval as the EBVs can be estimated early in life of an individual thus allowing breeders to select superior stock at a younger age (Meuwissen et al., 2001). It improves estimation of sex-limited traits, as it is able to source the information from the genotypes of both sexes, as opposed to phenotype-based estimations. Use of genotypes is also expected to improve predictions of breeding values for traits of low heritability, for which the identification of the genetic merit of an individual from phenotypic records and pedigree is difficult. It can also help alleviate pedigree errors, through accurate parental assignments (Visscher et al., 2002); finally in

some cases it could lead to reduced costs of collecting the phenotypic measurements (Goddard and Hayes, 2007).

3. USE OF MARKERS - MAS AND GAS

Since the first identification of markers, be it blood groups, or Single Nucleotide Polymorphisms, breeders have been trying to find useful associations between those markers and traits of economic importance in livestock species. There are three main ways in which that association may exist (Dekkers, 2004):

- direct effect of a marker on a trait in question, where the marker is the causative gene
- linkage disequlibrium (LD) between a marker and the causative gene
- Linkage between a marker and the causative gene, but present only within families.

The initial search for the direct information on the genes an individual carries was aimed at locating loci with some effect on a trait, referred to as quantitative trait loci (QTLs). The successfully localized QTLs could then be used in Gene Assisted Selection (GAS) approach, where a fixed effect of the QTL genotype was fitted in mixed linear models alongside with the standard polygenic effect estimated from the phenotypes and the pedigree (Kennedy et al., 1992). However, mapping of QTLs is frequently a challenging process, with a relatively large rate of false positives detected unless a validation study in a different population is used to confirm the results, or a stringent significance threshold is used (Belknap et al., 1996). For traits where QTL discovery and subsequent genotyping proves difficult or impossible, an alternative method called Marker Assisted Selection (MAS) was developed, in which marker genotypes were used as an approximation of the genotype at unknown QTL. The success of this approach depends largely on the relationship between the marker and the QTL, which can be described in terms of LD. The extent of LD is directly related to the recombination rate of given genomic region (Hartl and Clark, 1997), distance between loci and historical effective population size (Abasht et al., 2009). LD arises as an effect of three processes: admixture, genetic drift and selection; and is limited by the recombination (Lande and Thompson, 1990).

Compared to GAS, marker information used in MAS was used to estimate the marker variance, which accounted for some uncertainty in the prediction of QTL genotype from the markers. Because the exact location of the QTL was no longer required, the preliminary

research requirements for MAS were significantly reduced, as compared to GAS. MAS approach was also more versatile, with some markers being potentially linked to QTLs affecting different traits, thus allowing the use of marker data in evaluations of multiple traits (Dekkers, 2004). Both MAS and GAS faced similar problems in practical application in commercial setting. For most commercial traits, the variance observed is determined by many QTLs, therefore methods such as GAS and MAS, concentrating on one or several loci will only explain a small proportion of the total variance. This could be alleviated by fitting the marker terms alongside the polygenic term in the model, with the latter containing the proportion of variance not explained by the markers, however, this required large computational capacity, not readily available in 1990's. Other limitations included large numbers of markers needed for mapping, and large population sizes needed for accurate estimation of QTL effects, particularly for traits with low heritability (Lande and Thompson, 1990). In contrast to **A** matrix, for which an effective and simple and direct method of calculating the inverse has been developed (Henderson, 1976), there was no such method for calculating the inverse of the marker (co)variance matrix, which led to high computational demands. MAS faced also other challenges e.g. recombination breaking the associations between markers and QTLs, and problems with assessing the inheritance of QTL effect when parental markers were non-informative (Villanueva et al., 2002). Overall, the additional responses obtained by using genotypes in MAS and GAS strategies were found to decline over generations as a result of fixation at the QTL (Villanueva et al., 2005). While any form of selection leads to fixation of favourable alleles, this process is largely sped up in these two methodologies.

Despite these issues, the GAS method found use in commercial breeding in many livestock species, including beef and dairy cattle, poultry, pigs and sheep (Dekkers, 2004). The use of MAS has been less wide-spread, however this method too found its use in pigs, cattle and sheep (Dekkers, 2004). As the number of studies into the use of markers grew, it was realized that increasing the number of markers results in nearly additive increase in genetic gain (Haley and Visscher, 1998). This observation led to a revolutionary paper by Meuwissen *et al.* in 2001, in which markers spread across the genome were used to estimate the genome-wide breeding values (GEBVs). The method presented was termed as genomic selection (GS) and formed an innovative speculation as to what could be achieved with dense marker maps, which at the time were not available for any of the livestock species. GS presented a completely new approach to utilizing genotypes in predictions of breeding values, as although the interpretation of the results was still based on the linkage between

markers and QTLs, the search for actual positions and magnitudes of the effect of the causative loci were no longer considered as the main points of the experiment.

4. GENOMIC PREDICTION (GP)

In comparison to MAS, where individual markers or small groups of markers linked to QTLs were searched for by significance tests of their effects, genomic prediction (GP) is based on partition of the total variance through simultaneous estimation of all effects from all markers (Meuwissen et al., 2001). Furthermore, MAS at low marker densities requires that the linkage phase between marker and QTL has to be estimated separately in each family subject to selection (Meuwissen et al., 2001). Increasing marker densities allows establishing chromosome segments constituted of closely linked markers, more likely to be identical by descent (IBD), which removes the requirement of establishing linkage phase in separate families (Meuwissen et al., 2001). The IBD probability between any two alleles at a given locus represents the correlation between unknown effects of these alleles (Meuwissen, 2003). However, there is a large number of these segments in the genome, which requires estimation of large number of effects, usually larger than the number of phenotypic points from which they are to be estimated (Meuwissen et al., 2001). In addition, these segments are not constant entities, with recombination breaking up the associations from generation to generation. This poses a potential problem in analyses, where the number of effects to be estimated (p) far exceeds the number of records (n) available for the estimation ($p>>n$), causing shortages in the degrees of freedom in the analyses. Since 2001, numerous methodologies have been thus proposed to cope with this and other challenges of GP.

5. METHODS OF GENOMIC PREDICTION (GP)

The initial methods of GP attempted to limit the discrepancy between the numbers of effects to estimate and available observations by including in the analyses only markers with observed effect on a trait. One of the earliest approaches, the Least Squares (LS) method tackled this problem by stepwise addition of genes which effects exceeded some significance threshold, with remaining effects of non-significant loci set to 0 (Meuwissen et al., 2001). However, the latter seems to be unfair, as the distribution of gene effects appears to be more leptokurtic than normal, with occasional loci of large effect (Hayes and Goddard, 2001). As such, a proportion of the genetic variance is likely to be explained by multiple loci of small effect, for which an individual locus effect is too small to pass a formal significance

threshold. A small effect does not necessarily mean lack of effect, therefore Least Squares method is potentially discarding large amounts of useful data. The proportion of the variance which could be explained by these removed markers is then wrongly attributed to the few loci that passed the statistical threshold, resulting in the overestimation of their effects (Meuwissen et al., 2001).

An alternative approach to GP which deals with the shortage in degrees of freedom is based on Bayesian statistics. There is a number of methods developed under the Bayes Theorem with different underlying assumptions, however all of them are based on the same framework. In Bayesian methods applied to GP, the marker effects are fitted as random, which means that they do not require the assignment of degrees of freedom. The markers can be categorized according to their effect size, with the number of markers with major effects determined by a prior (Meuwissen, 2003). The marker variances typically follow inverted chi-square distribution, while the distribution of marker effects varies between particular versions of Bayesian approach. In contrast to LS, the categorization of marker effects in Bayesian methods is carried out simultaneously with the estimation of the variance and breeding values, thus making this approach dynamic.

Although Bayesian methods have been found to perform well in multiple simulation studies (e.g. Meuwissen et al., 2001, Goddard and Hayes, 2007, Habier et al., 2007), these methods perform best when there are few QTLs affecting a trait (Daetwyler et al., 2010). In scenarios where a trait of interest is determined by multiple loci, a modification of the traditional BLUP to incorporate marker information has been found to be almost as accurate (e.g. Yang et al., 2010, VanRaden, 2008, Clark et al., 2011) or even better than Bayesian methods (Daetwyler et al., 2010). Genomic BLUP (GBLUP) differs from the classical BLUP in the relationship matrix used, i.e. the **A** of BLUP is replaced with genomic relationship matrix (**G**) calculated from the marker genotypes (Habier et al., 2013). GBLUP deals with shortage of degrees of freedom in similar way as Bayesian methods, i.e. by fitting marker effects as random. However, in contrast to most Bayes implementations, GBLUP assumes a constant variance across loci, i.e. all loci are assumed to contribute equally to the total variance (Meuwissen and Goddard, 2001). This approach resembles the ridge regression methodology. In ridge regression, the regression coefficients of the marker scores are subject to shrinkage, with the shrinkage parameter λ calculated for the given dataset in such way, as to minimize the model error (Whittaker et al., 2000). In GBLUP approach, the shrinkage occurs in the calculation of the genomic relationships, with the genotypes of the individuals

centred and standardized, and the λ used being a simple noise-to-signal ratio of residual and additive variances (de los Campos et al., 2012).

6. GENOMIC RELATIONSHIP MATRICES

The role of relationship matrices in BLUP approaches is to approximate the genetic resemblance between relatives at the locus of interest. Traditionally it has been done on the basis of pedigree. By using marker genotypes, IBD probabilities can be calculated for the marker positions from which the QTL relationships are derived. The reliability of this extrapolation depends on the LD between the marker and unknown QTL (Habier et al., 2013). In the best-case scenario, the two loci are in complete LD, so the prediction of QTL genotype based on marker genotype is estimated with high confidence and remains unchanged until recombination breaks the association. In cases where the LD is not complete, using a marker-derived relationship matrix may result in loss of the proportion of the genetic variance captured. However, marker genotypes can explain the relationships at QTL loci irrespective of LD when the individuals are related through pedigree (de los Campos et al., 2013). In such cases, the two loci are likely to co-segregate, thus providing the information on QTLs from marker genotypes. As commercial livestock populations frequently consist of related animals, a linkage analysis approach (LA) which utilizes pedigree relationships could be an attractive alternative when the amount of LD between markers and QTLs is in question (Meuwissen and Goddard, 2010).

7. ACCURACY AND BIAS

Irrespective of the methodology used and rationale behind it, the practical utility of genomic prediction is usually evaluated through the estimates of bias and accuracy. Ideally, the performance of (G)EBV prediction should be estimated in relation to the true breeding value (TBV), as has been done in studies based on simulations where the latter is known (e.g. Meuwissen et al., 2001). However, in real life datasets the TBV cannot be estimated without an error. Therefore, the performance of the methods can be evaluated by comparing the predicted (G)EBVs to the phenotypic records of a selection candidate (Wolc et al., 2010). Cross-validation is one of the most often used approaches for calculation of the accuracy of GEBVs. In cross-validation, the data is divided into subsets: training (TRN) and testing (TST). The TRN includes individuals with both genotypes and phenotypic records, on the basis of which the marker effects are estimated, while the TST includes selection candidates

(Goddard and Hayes, 2007). The most stringent test of the accuracy is obtained when the phenotypes of the selection candidates are masked for the prediction analysis, and the GEBV is based on the genotype of the candidate only.

Bias of prediction is obtained as a regression coefficient (β) of phenotype or TBV on the predicted (G)EBV. A property of BLUP methodology is that it is unbiased, with the average deviation of (G)EBV from the TBV expected to be 0 (Henderson, 1973). When this condition holds, β=1. Departures from this value indicate differences in the distribution of the TBVs and (G)EBVs, i.e. β>1 indicates that the range of TBVs exceeds the range of EBVs, while β<1 indicates on overestimation of the TBV range. The latter is usually caused by selection. As long as the data used in the process of selection is incorporated in the evaluation, traditional BLUP can cope with this nuisance by analysing the distribution conditional on the selected variable (Henderson, 1973). This method however is also based on an assumption that the additive genetic covariance matrix ($\boldsymbol{A}\sigma_a^2$, where σ_a^2 is the additive genetic variance) is estimated without error. The original proofs for this calculation have been based on the numerator relationship matrix, therefore the efficacy of GBLUP in dealing with selection bias can potentially depart from this expectation.

The accuracy of breeding value predictions using marker data is limited by two main factors (Meuwissen et al., 2001):

a) Incomplete LD – where QTL variance is not fully explained by the markers (Meuwissen et al., 2001).

b) Sampling error (co)variances for the estimation of the marker effects – which is broadly determined by environmental variance divided by number of animals, and thus depends on the number of genotyped and recorded animals and heritability of the trait in question. Hence, for a given accuracy, the number of records for a trait with lower heritability needs to be increased (Meuwissen et al., 2001)

Thus, the accuracy of the (G)EBV prediction depends on the characteristics of the trait (heritability, distribution of marker effects), population (effective population size), study design (number of genotyped and phenotyped individuals, number of markers used for genotyping) and the genome itself (length of the genome, number of independent chromosome segments and recombination rates) (Daetwyler et al., 2008). Compared to other livestock species, the genome characteristics of chickens are distinctly different.

8. CHICKEN GENOME

The chicken was the first livestock species to have its genome sequenced, with the first draft published in 2004 by International Chicken Genome Sequencing Consortium (Hillier et al., 2004). The chicken genome is only a third of mammalian genome size at about 1.05Gb, organised over 38 autosomes (2n=78) and 2 sex chromosomes, Z and W, with female being a heterogametic sex. The chicken chromosomes vary considerably in their physical lengths, from nearly 200Mb for chromosome 1 to less than 2Mb for chromosome 25, with the currently unmapped chromosomes 29-38 likely to have even shorter lengths (Groenen et al., 2009). Due to this large variation in size, chicken chromosomes have been divided into macrochromosomes (chromosomes 1 – 5), intermediate (chromosomes 6 – 10) and microchromosomes (chromosomes 11 – 38) (Hillier et al., 2004). Although current sequencing data covers approximately 96% of the chicken genome, the contigs of the microchromosomes are notoriously difficult to map (Groenen et al., 2009, Hillier et al., 2004, Rubin et al., 2010, Kranis et al., 2013). It appears that some microchromosome sequences are not cloned properly by *E. coli* in the process of construction of the BAC libraries, possibly due to high concentration of GC content (Hillier et al., 2004).

The physical size of the chromosomes in chickens was found to be negatively correlated to recombination rate (cM/Mb), GC content, CpG island density, gene density, gene length and repeat density (Hillier et al., 2004, Groenen et al., 2009). The overall recombination rates of chicken genome range between 2.5 and 21cM/Mb, depending not only on the chromosome, but also on the particular parts of a chromosome (Ellegren, 2010). A difference in recombination rates has been also observed between different chicken populations, with some estimates reaching as high as 74cM/Mb for some genomic regions in a broiler dam line (Elferink et al., 2010). These recombination rates far exceed the recombination rates observed in mammals (1 – 2cM/Mb in humans, 0.5 – 1.0cM/Mb in mouse) (Hillier et al., 2004, Ellegren, 2010).

The recombination patterns between parts of a genome are represented through linkage maps. The first map of chicken genome was constructed in 1992 by Bumstead and Palyga. This map was based on Restricted Fragment Length Polymorphisms (RFLPs), and included 100 markers spread over the genome (Bumstead and Palyga, 1992). Since then, the linkage map has constantly been improved, with numerous versions, and change of the marker types. In 2000, Groenen et al. published the first Consensus Linkage Map of the Chicken Genome. This map was based mainly on microsatellite markers, which formed 50 linkage groups with

a total of 1,889 loci. Using these markers, the length of the linkage map of chicken genome was estimated at 4,000cM (Groenen et al., 2000).

In 2009, an improved version of linkage map for chicken was released, including nearly 9K markers, predominantly SNPs (Groenen et al., 2009). This map estimated the length of sex averaged linkage map at 3,228cM. The decrease in estimated length compared to previous maps was attributed to reduced error rate achieved through the type of markers used and their density. Authors of this study observed also a difference in linkage map lengths between populations, distributed over the genome rather than located at one region. The map identified 34 linkage groups, representing at least 29 chromosomes. However, even with large number of markers used, there are still at least 5 microchromosomes not represented in any of the available maps.

Since first maps of the chicken genome were constructed, researchers tried to estimate the extent and distribution of the LD in chicken. Livestock species are thought to have higher levels of LD than human populations, as the forces generating LD (genetic drift, admixture, selection) are more frequently observed in animal populations (Rao et al., 2008). It has been suggested that short blocks of LD represent distant ancestral populations and as such are regulated by not only recombination but also mutation and drift (Hayes et al., 2003). Larger blocks of LD are believed to represent more recent generations, and as such are not likely to be affected by mutation or drift, whose effect in initial generations is limited (Falconer and Mackay, 1996).

Considering the dependence of the extent of LD blocks on recombination rates, it follows that the distribution of such blocks will vary between populations and genomic regions. Indeed, differences between breeds and lines of chickens have been documented (e.g. Muir et al., 2008b, Aerts et al., 2007). Overall, layer chickens, e.g. White Leghorns, were found to have the largest blocks of LD accompanied by the lowest levels of heterozygosity, while a population of broiler chickens was found to have shortest blocks of LD out of the populations studied, with blocks rarely extending over 0.5cM (Andreescu et al., 2007). The extent of LD was also found to be larger on macrochromosomes, which was expected because of the negative correlation between physical chromosome size and recombination rates (Megens et al., 2009).

Due to the complex history of chickens, starting from natural selection of wild ancestors, through domestication, to breed formation and intense artificial selection, the chicken genome was and is subject to numerous processes prior, during and post domestication

(Sundstrom et al., 2004). The effect of the artificial selection and consequential reduction in genetic variance as a result of increased inbreeding rates is a topic of controversy. It has been suggested that the genetic diversity of modern chickens may be severely compromised due to a limited number of lines from which commercial lines originated, specialized breed utilization, intensive selection and breeding structure, leading to inbreeding (Muir et al., 2008a). This speculation was initially formed based on the results of a study on over 2,500 birds genotyped for 3K SNP markers, which found that broiler and layer lines had reduced genetic diversity when compared to non-commercial birds, such as Red Jungle Fowl, Chinese Silkie or some experimental lines (Muir et al., 2008a). By calculating a proportion of SNP alleles missing in the commercial birds, as compared to a Hypothetical Ancestral Population - HAP, the authors concluded that the modern commercial breeding lines have lost 60% or more of the genetic diversity (Muir et al., 2008a). A similar result was reported by Abasht et al. (2009) who found that only 35% of genotyped SNPs were segregating in studied two lines of layer chickens. This idea was subsequently supported by studies based on limited numbers of lines and markers used (predominantly microsatellite markers), which found reduced levels of polymorphisms in commercial breeds and lines, as compared to wild ancestors and native African and Asian breeds (as reviewed by Groeneveld et al., 2010). However, recent analyses of the chicken genome using sequence data and large number of chicken lines revealed around 78 million segregating SNPs in one or more chicken lines, with average numbers of segregating SNPs per line ranging between 448K and 11M, pooled to the average of 10M SNPs in broilers, 7.7M in layers and 1M in experimental inbred lines (Kranis et al., 2013). The same study showed that out of 10M SNPs that passed different quality criteria checks, around 23% were common to 24 lines including broiler, layer and inbred lines, and over 31% were common to 15 lines including broiler and layer lines (Kranis et al., 2013). This finding contradicts the theory of loss of genomic diversity due to domestication of a small pool of individuals, which later gave rise to current chicken population worldwide. Similar results were found by Bovine HapMap Consortium (Gibbs et al., 2009), where despite decline in effective population size of bovines due to domestication, breed formation and artificial selection, nucleotide diversity calculated both for taurine and indicine breeds is still very high (40% higher than in humans).

The effects of domestication and artificial selection affect also the effective population size. Effective population size, N_e, is a parameter used to describe how many randomly mating individuals under random selection would be needed to produce a population of offspring with the same rate of inbreeding, as observed in the studied population (Falconer and Mackay, 1996). It follows that heavily selected livestock populations will have a reduced

estimate of N_e. The reduction in N_e in livestock has been confirmed in several studies, e.g. in cattle (Gibbs et al., 2009), and stands in contrast with apparently increasing estimates of N_e in human populations (Tenesa et al., 2003) The estimates of N_e in chicken differ between populations, i.e. in a study on two elite lines of layers the N_e were estimated below 30 (Abasht et al., 2009), while N_e of broilers was reported between 50 and 200, depending on line (Andreescu et al., 2007).

Irrespective of the possible reduction of genetic variance observed in modern chicken populations, the advances in genotyping and sequencing methodologies result in continuous discovery of causative loci. At the time of writing this Chapter, there were 4,282 QTLs identified in the Chicken QTL database, affecting 305 different traits (Chicken QTLdb, 2014).

9. GENOMIC SELECTION IN CHICKENS

Application of GS in poultry was delayed compared to other livestock species, e.g. cattle, which had their genome sequenced later than chicken and yet are subject to routine GS on a wide scale (Hayes et al., 2009a). There are several factors which are often used to explain this delay. Firstly, the major benefit of GS in reducing the generation interval in mammalian species offers little profit for poultry breeders, particularly broilers, as most of the important traits in broiler selection candidates can be measured by puberty, thus not affecting the generation interval (Garrick, 2010). Secondly, implementation of the GS on a commercial scale presents large monetary challenge, both in the preliminary set up (i.e. development of genotyping platform, establishing necessary parameters for successful implementation of the method between lines) and throughout the breeding program, with the costs of genotyping of selection candidates far exceeding the monetary value of the animal (Avendano et al., 2010; Preisinger, 2012). Thirdly, separation of poultry production into broilers and layers, and subsequent division into highly specialized lines with specific breeding goals may lead to multiple performance testing analyses for a GS method (Preisinger, 2012). Finally, the structure of the breeding programme of chickens, with highly hierarchical structure and extensive use of heterosis complicates the expectation of the benefits that may be obtained on commercial farms (Preisinger, 2012).

Despite these perceived drawbacks, the application of GP in poultry has been gaining momentum in the last decade. The first genotyping chip for poultry followed soon after the first draft of the chicken genome was published and it was a proprietary chip developed by

Aviagen Ltd in 2005 (Avendano et al., 2010). Since then, the work continued, followed by development of a 3K Illumina chip (Muir et al., 2008b), a 42K Illumina iSelect BeadChip, which was the property of EW Group (Avendano et al., 2010), and a 60K Illumina SNP BeadChip (Groenen et al., 2011). The most recent and extensive project resulted in the creation of a 600K Affymetrix Axiom genotyping array, covering larger phylogeny than any of the predecessors, with 25 lines representing not only layers and broilers, but also experimental lines (Kranis et al., 2013). This high density chip is the first one to be commercially available to the general public.

Due to fairly recent development of the genotyping chips and the proprietary nature of most of the arrays, the efficiency and benefits of GS in chicken production is still largely unknown. The initial studies used variable methods, e.g. machine learning (Long et al., 2007), linear models using non-parametrically derived kernels (Gonzalez-Recio et al., 2008), Bayesian methods (e.g. Wolc et al., 2011) and GBLUP (e.g. Andreescu et al., 2010). Most of the published studies showed some benefits of using marker information, as shown in the examples presented in Table 1.1, with the ranking between methods depending on the study design. e.g. size of the analysed population, genetic architecture of the traits etc. Due to the short time in which genotyping arrays have been available, most of the published studies were carried out on a limited numbers of birds, leading to large standard errors of the estimates, which in turn reduce the informativeness of the results. Similarly, until recently, the studies were limited in terms of the marker densities.

10. THE AIM OF THIS STUDY AND THESIS OUTLINE

The purpose of this study was to evaluate the efficacy of the GBLUP methodology in prediction of GEBVs for a commercial population of broiler chickens. The evaluations were carried out on unprecedented scale, with the size of population studied exceeding previously reported, and the density of markers greater than previously utilized in poultry. The assessment of genomic predictions was preceded by analyses based on pedigree, which provided an overview of the genetic variance present in the commercial broiler populations.

Chapter 2 presents an examination of the pedigree-based REML models applied to 6 key traits. Models tested include several components of the variance, including maternal genetic effects which are not used in routine commercial evaluations. The analyses performed provide an up to date estimates of the heritability and components of the trait variance.

Chapter 3 is also based on traditional REML analysis, applied to two key broiler traits. In this Chapter, a vast dataset including 1.3 million birds spread over 24 generations was used for an analysis of the changes in variance with time. It also provides an estimate of the accuracy and bias of EBV predictions obtained from pedigree-based BLUP.

Chapter 4 provides preliminary estimates of accuracy of GBLUP on a moderately sized population of broilers (over 5,400 individuals), using chips of differing densities, reaching up to over 431K SNPs. The traits analysed in this Chapter were sex limited and fitness traits, which are difficult to measure and hence could benefit most from the use of markers.

Chapter 5 investigates the use of different genomic relationship matrices on the accuracy and bias of GBLUP predictions. The matrices were constructed using LD and linkage analysis (LA) approach, thus changing the population to which the estimates of variance refer to. The matrices were also created by regressing the genomic relationships (LD) back to the pedigree based relationships, either sourced from **A**, or matrix based on LA approach, using different regression coefficients. In addition, the analyses presented in this Chapter compare the effect of marker choice, by using an evenly spaced marker chip and a chip constructed on markers selected through GWAS.

Chapter 6 presents the comprehensive evaluation of the utility of GBLUP applied to broiler production by using a large population of birds (over 23,500 genotyped and phenotyped individuals) genotyped with several chip densities, reaching up to 412K SNPs. Thanks to the large number of individuals used, the Chapter provides information on the effect of the number of individuals used for evaluations. The empirical results obtained in this Chapter are then compared to the theoretical predictions of accuracy.

The work presented shows a thorough analysis of the possible modifications to the genetic evaluations in chicken and, thanks to the large amounts of data available, add to the general knowledge of the necessary population and trait parameters needed for successful genetic selection programmes.

Table 1.1 Examples of studies into GP using real data on chicken populations, with the benefit of the genomic over the pedigree-based methods given in the last column. The NA qualifier is given when no reference value was provided in the study.

Study	Number of genotyped individuals	Marker Density	Method	Trait	Benefit of GP
Long et al. (2007)	200	5K SNP	Machine Learning	Early mortality	NA
Gonzalez-Recio et al. (2008)	200	5K SNP	- Linear Regression - Kernel Regression - RKHS regression	Late mortality	up to 100%
González-Recio et al. (2009)	394	4K SNP	- Bayes A - RKHS regression	Food conversion	145%
Andreescu et al. (2010)	Not Given	12K SNP	- GBLUP - Bayes B - Bayes C - Bayes C-π	Juvenile body weight	NA
Chen et al. (2011)	3,284 & 3,098	57K SNP	Single step GBLUP (ssGBLUP)	- Body weight - Breast mass - Leg score	Up to 70%
Wolc et al. (2011)	2,708	26K SNP	- GBLUP - Bayes C-π	Range of egg production traits	Up to 88%
Simeone et al. (2012)	3,195 & 3,001	57K SNP	ssGBLUP	Juvenile body weight	NA
Abdollahi-Arpanahi et al. (2014)	1,351	580K	GBLUP	Juvenile body weight, Breast mass and Hen housed production	NA

11. REFERENCES

ABASHT, B., SANDFORD, E., ARANGO, J., SETTAR, P., FULTON, J. E., O'SULLIVAN, N. P., HASSEN, A., HABIER, D., FERNANDO, R. L., DEKKERS, J. C. & LAMONT, S. J. 2009. Extent and consistency of linkage disequilibrium and identification of DNA markers for production and egg quality traits in commercial layer chicken populations. *BMC Genomics,* 10 Suppl 2**,** S2.

ABDOLLAHI-ARPANAHI, R., NEJATI-JAVAREMI, A., PAKDEL, A., MORADI-SHAHRBABAK, M., MOROTA, G., VALENTE, B., KRANIS, A., ROSA, G. & GIANOLA, D. 2014. Effect of allele frequencies, effect sizes and number of markers on prediction of quantitative traits in chickens. *Journal of Animal Breeding and Genetics*,131, 123 - 133

AERTS, J., MEGENS, H. J., VEENENDAAL, T., OVCHARENKO, I., CROOIJMANS, R., GORDON, L., STUBBS, L. & GROENEN, M. 2007. Extent of linkage disequilibrium in chicken. *Cytogenet Genome Res,* 117**,** 338-45.

ANDREESCU, C., AVENDANO, S., BROWN, S. R., HASSEN, A., LAMONT, S. J. & DEKKERS, J. C. 2007. Linkage disequilibrium in related breeding lines of chickens. *Genetics,* 177**,** 2161-9.

ANDREESCU, C., HABIER, D., FERNANDO, R., KRANIS, A., WATSON, K., AVENDANO, S. & DEKKERS, J. C. M. Accuracy of Genomic Predictions Across Breeding Lines of Chickens. World Congress on Genetics Applied to Livestock Production, 2010 Leipzig.

AVENDANO, S., WATSON, K. & KRANIS, A. Genomics in Poultry Breeding - From Utopias to Deliverables. World Congress on Genetics Applied to Livestock Production, 2010 Leipzig.

BELKNAP, J., MITCHELL, S., O'TOOLE, L., HELMS, M. & CRABBE, J. 1996. Type I and type II error rates for quantitative trait loci (QTL) mapping studies using recombinant inbred mouse strains. *Behavior genetics,* 26**,** 149-160.

BUMSTEAD, N. & PALYGA, J. 1992. A preliminary linkage map of the chicken genome. *Genomics,* 13**,** 690-697.

CHEN, C. Y., MISZTAL, I., AGUILAR, I., TSURUTA, S., MEUWISSEN, T. H. E., AGGREY, S. E., WING, T. & MUIR, W. M. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. *J Anim Sci,* 89**,** 23-28.

CLARK, S., HICKEY, J. & VAN DER WERF, J. 2011. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution,* 43**,** 18.

DAETWYLER, H. D., PONG-WONG, R., VILLANUEVA, B. & WOOLLIAMS, J. A. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics,* 185**,** 1021-31.

DAETWYLER, H. D., VILLANUEVA, B. & WOOLLIAMS, J. A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One,* 3**,** e3395.

DE LOS CAMPOS, G., HICKEY, J. M., PONG-WONG, R., DAETWYLER, H. D. & CALUS, M. P. L. 2012. Whole Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics*, 193, 327 - 345

DE LOS CAMPOS, G., VAZQUEZ, A. I., FERNANDO, R., KLIMENTIDIS, Y. C. & SORENSEN, D. 2013. Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genet,* 9**,** e1003608.

DEKKERS, J. C. M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *Journal of Animal Science,* 82**,** E313-E328.

DUNNINGTON, E. A., HONAKER, C. F., MCGILLIARD, M. L. & SIEGEL, P. B. 2013. Phenotypic responses of chickens to long-term, bidirectional selection for juvenile body weight—Historical perspective. *Poultry Science,* 92**,** 1724-1734.

ELFERINK, M., VAN AS, P., VEENENDAAL, T., CROOIJMANS, R. & GROENEN, M. 2010. Regional differences in recombination hotspots between two chicken populations. *BMC Genetics,* 11**,** 11.

ELLEGREN, H. 2010. Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution,* 25**,** 283 - 291.

FALCONER, D. S. & MACKAY, T. F. C. 1996. *Introduction to quantitative genetics / D.S. Falconer and Trudy F.C. Mackay,* Harlow : Longman, 1996. 4th ed.

FAO, F. A. A. O. O. T. U. N. 2014. *FAO Food Statistics* [Online]. Available: http://faostat3.fao.org/faostat-gateway/go/to/download/FB/FBS/E.

GARRICK, D. 2010. The revolution in genetic improvement. *Lohmann Information,* 45**,** 9 - 17.

GIBBS, R. A., TAYLOR, J. F., VAN TASSELL, C. P., BARENDSE, W., EVERSOLE, K. A., GILL, C. A., GREEN, R. D., HAMERNIK, D. L., KAPPES, S. M., LIEN, S., MATUKUMALLI, L. K., MCEWAN, J. C., NAZARETH, L. V., SCHNABEL, R. D., WEINSTOCK, G. M., WHEELER, D. A., AJMONE-MARSAN, P., BOETTCHER, P. J., CAETANO, A. R., GARCIA, J. F., HANOTTE, O., MARIANI, P., SKOW, L. C., SONSTEGARD, T. S., WILLIAMS, J. L., DIALLO, B., HAILEMARIAM, L., MARTINEZ, M. L., MORRIS, C. A., SILVA, L. O., SPELMAN, R. J., MULATU, W., ZHAO, K., ABBEY, C. A., AGABA, M., ARAUJO, F. R., BUNCH, R. J., BURTON, J., GORNI, C., OLIVIER, H., HARRISON, B. E., LUFF, B., MACHADO, M. A., MWAKAYA, J., PLASTOW, G., SIM, W., SMITH, T., THOMAS, M. B., VALENTINI, A., WILLIAMS, P., WOMACK, J., WOOLLIAMS, J. A., LIU, Y., QIN, X., WORLEY, K. C., GAO, C., JIANG, H., MOORE, S. S., REN, Y., SONG, X. Z., BUSTAMANTE, C. D., HERNANDEZ, R. D., MUZNY, D. M., PATIL, S., SAN LUCAS, A., FU, Q., KENT, M. P., VEGA, R., MATUKUMALLI, A., MCWILLIAM, S., SCLEP, G., BRYC, K., CHOI, J., GAO, H., GREFENSTETTE, J. J., MURDOCH, B., STELLA, A., VILLA-ANGULO, R., WRIGHT, M., AERTS, J., JANN, O., NEGRINI, R., GODDARD, M. E., HAYES, B. J., BRADLEY, D. G., BARBOSA DA SILVA, M., LAU, L. P., LIU, G. E., LYNN, D. J., PANZITTA, F. & DODDS, K. G. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science,* 324**,** 528-32.

GODDARD, M. E. & HAYES, B. 2007. Genomic Selection. *Journal of Animal Breeding and Genetics,* 124**,** 323 - 330.

GODDARD, M. E., HAYES, B. & MEUWISSEN, T. Genomic Selection in Farm Animal Species - Lessons Learnt and Future Perspectives. World Congress on Genetics Applied to Livestock Production, 1 - 6 August 2010 2010 Leipzig.

GONZALEZ-RECIO, O., GIANOLA, D., LONG, N., WEIGEL, K. A., ROSA, G. J. M. & AVENDANO, S. 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. *Genetics,* 178**,** 2305-2313.

GONZÁLEZ-RECIO, O., GIANOLA, D., ROSA, G. J., WEIGEL, K. A. & KRANIS, A. 2009. Genome-assisted prediction of a quantitative trait measured in parents and progeny: application to food conversion rate in chickens. *Genet. Sel. Evol,* 41 (3)

GROENEN, M., MEGENS, H.-J., ZARE, Y., WARREN, W., HILLIER, L., CROOIJMANS, R., VEREIJKEN, A., OKIMOTO, R., MUIR, W. & CHENG, H. 2011. The development and characterization of a 60K SNP chip for chicken. *BMC Genomics,* 12**,** 274.

GROENEN, M. A., CHENG, H. H., BUMSTEAD, N., BENKEL, B. F., BRILES, W. E., BURKE, T., BURT, D. W., CRITTENDEN, L. B., DODGSON, J., HILLEL, J., LAMONT, S., DE LEON, A. P., SOLLER, M., TAKAHASHI, H. & VIGNAL, A. 2000. A consensus linkage map of the chicken genome. *Genome Res,* 10**,** 137-47.

GROENEN, M. A., WAHLBERG, P., FOGLIO, M., CHENG, H. H., MEGENS, H. J., CROOIJMANS, R. P., BESNIER, F., LATHROP, M., MUIR, W. M., WONG, G. K., GUT, I. & ANDERSSON, L. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res,* 19**,** 510-9.

GROENEVELD, L., LENSTRA, J., EDING, H., TORO, M., SCHERF, B., PILLING, D., NEGRINI, R., FINLAY, E., JIANLIN, H. & GROENEVELD, E. 2010. Genetic diversity in farm animals–a review. *Animal Genetics,* 41**,** 6-31.

HABIER, D., FERNANDO, R. L. & DEKKERS, J. C. M. 2007. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics,* 177**,** 2389-2397.

HABIER, D., FERNANDO, R. L. & GARRICK, D. J. 2013. Genomic-BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics,* 194, 597 - 607

HALEY, C. S. & VISSCHER, P. 1998. Strategies to Utilize Marker - Quantitative Trait Loci Associations. *Journal of Dairy Science,* 81**,** 85 - 97.

HARTL, D. L. & CLARK, A. G. 1997. *Principles of population genetics,* Sunderland, MA, Sinauer Associates.

HAVENSTEIN, G., FERKET, P. & QURESHI, M. 2003. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poult Sci,* 82**,** 1500-1508.

HAYES, B. J., BOWMAN, P. J., CHAMBERLAIN, A. J. & GODDARD, M. E. 2009a. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science,* 92**,** 433-443.

HAYES, B. J., VISSCHER, P. M. & GODDARD, M. E. 2009b. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research,* 91**,** 47-60.

HAYES, B. J., VISSCHER, P. M., MCPARTLAN, H. C. & GODDARD, M. E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res,* 13**,** 635-43.

HENDERSON, C. R. Estimation of genetic parameters. *The Annals of Mathematical Statistics,* 21, 309

HENDERSON, C. R. 1973. Sire evaluation and genetic trends. *Journal of Animal Science,* 1973**,** 10-41.

HENDERSON, C. R. 1976. A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics,* 32**,** 69-83.

HILLIER, L. W., MILLER, W., BIRNEY, E., WARREN, W., HARDISON, R. C., PONTING, C. P., BORK, P., BURT, D. W., GROENEN, M. A. M., DELANY, M. E., DODGSON, J. B., CHINWALLA, A. T., CLIFTEN, P. F., CLIFTON, S. W., DELEHAUNTY, K. D., FRONICK, C., FULTON, R. S., GRAVES, T. A., KREMITZKI, C., LAYMAN, D., MAGRINI, V., MCPHERSON, J. D., MINER, T. L., MINX, P., NASH, W. E., NHAN, M. N., NELSON, J. O., ODDY, L. G., POHL, C. S., RANDALL-MAHER, J., SMITH, S. M., WALLIS, J. W., YANG, S. P., ROMANOV, M. N., RONDELLI, C. M., PATON, B., SMITH, J., MORRICE, D., DANIELS, L., TEMPEST, H. G., ROBERTSON, L., MASABANDA, J. S., GRIFFIN, D. K., VIGNAL, A., FILLON, V., JACOBBSON, L., KERJE, S., ANDERSSON, L., CROOIJMANS, R. P. M., AERTS, J., VAN DER POEL, J. J., ELLEGREN, H., CALDWELL, R. B., HUBBARD, S. J., GRAFHAM, D. V., KIERZEK, A. M., MCLAREN, S. R., OVERTON, I. M., ARAKAWA, H., BEATTIE, K. J., BEZZUBOV, Y., BOARDMAN, P. E., BONFIELD, J. K.,

CRONING, M. D. R., DAVIES, R. M., FRANCIS, M. D., HUMPHRAY, S. J., SCOTT, C. E., TAYLOR, R. G., TICKLE, C., BROWN, W. R. A., ROGERS, J., BUERSTEDDE, J. M., WILSON, S. A., STUBBS, L., OVCHARENKO, I., GORDON, L., LUCAS, S., MILLER, M. M., INOKO, H., SHIINA, T., KAUFMAN, J., SALOMONSEN, J., SKJOEDT, K., WONG, G. K. S., WANG, J., LIU, B., WANG, J., YU, J., YANG, H. M., NEFEDOV, M., KORIABINE, M., DEJONG, P. J., GOODSTADT, L., WEBBER, C., DICKENS, N. J., LETUNIC, I., SUYAMA, M., TORRENTS, D., VON MERING, C., et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature,* 432**,** 695-716.

KENNEDY, B., QUINTON, M. & VAN ARENDONK, J. 1992. Estimation of effects of single genes on quantitative traits. *Journal of Animal Science,* 70**,** 2000-2012.

KRANIS, A., GHEYAS, A., BOSCHIERO, C., TURNER, F., YU, L., SMITH, S., TALBOT, R., PIRANI, A., BREW, F., KAISER, P., HOCKING, P., FIFE, M., SALMON, N., FULTON, J., STROM, T., HABERER, G., WEIGEND, S., PREISINGER, R., GHOLAMI, M., QANBARI, S., SIMIANER, H., WATSON, K., WOOLLIAMS, J. & BURT, D. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics,* 14**,** 59.

LANDE, R. & THOMPSON, R. 1990. Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genetics,* 124**,** 743-756.

LONG, N., GIANOLA, D., ROSA, G., WEIGEL, K. & AVENDANO, S. 2007. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. *Journal of Animal Breeding and Genetics,* 124**,** 377-389.

MEGENS, H.-J., CROOIJMANS, R., BASTIAANSEN, J., KERSTENS, H., COSTER, A., JALVING, R., VEREIJKEN, A., SILVA, P., MUIR, W., CHENG, H., HANOTTE, O. & GROENEN, M. 2009. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. *BMC Genetics,* 10**,** 86.

MEUWISSEN, T. Genomic Selection: The Future of Marker Assisted Selection and Animal Breeding. FAO International Workshop "Marker Assisted Selection: A Fast Track to Increase in Genetic Gain in Plant and Animal Breeding?"q, 2003 Turin.

MEUWISSEN, T. & GODDARD, M. 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution,* 33**,** 605 - 634.

MEUWISSEN, T. & GODDARD, M. 2010. The Use of Family Relationships and Linkage Disequilibrium to Impute Phase and Missing Genotypes in Up to Whole-Genome Sequence Density Genotypic Data. *Genetics,* 185**,** 1441-1449.

MEUWISSEN, T., HAYES, B. & GODDARD, M. E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics,* 157**,** 1819 - 1829.

MRODE, R. A. 2005. *Linear models for the prediction of animal breeding values*, Wallingford: CABI

MUIR, W. M., WONG, G. K., ZHANG, Y., WANG, J., GROENEN, M. A., CROOIJMANS, R. P., MEGENS, H. J., ZHANG, H., OKIMOTO, R., VEREIJKEN, A., JUNGERIUS, A., ALBERS, G. A., LAWLEY, C. T., DELANY, M. E., MACEACHERN, S. & CHENG, H. H. 2008a. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proc Natl Acad Sci U S A,* 105**,** 17312-7.

MUIR, W. M., WONG, G. K., ZHANG, Y., WANG, J., GROENEN, M. A. M., CROOIJMANS, R. P. M. A., MEGENS, H. J., ZHANG, H. M., MCKAY, J. C., MCLEOD, S., OKIMOTO, R., FULTON, J. E., SETTAR, P., O'SULLIVAN, N. P., VEREIJKEN, A., JUNGERIUS-RATTINK, A., ALBERS, G. A. A., LAWLEY, C. T., DELANY, M. E. & CHENG, H. H. 2008b. Review of the initial validation and characterization of a 3K chicken SNP array. *Worlds Poultry Science Journal,* 64**,** 219-225.

PREISINGER, R. 2012. Genome-wide selection in poultry. *Animal Production Science,* 52**,** 121-125.
QTLDB, C. *Chicken QTLdb* [Online]. Available: http://www.animalgenome.org/cgi-bin/QTLdb/GG/summary [Accessed 16/06/2014.
RAO, Y. S., LIANG, Y., XIA, M. N., SHEN, X., DU, Y. J., LUO, C. G., NIE, Q. H., ZENG, H. & ZHANG, X. Q. 2008. Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas,* 145**,** 251-7.
RUBIN, C.-J., ZODY, M., ERIKSSON, J., MEADOWS, J., SHERWOOD, E., WEBSTER, M., JIANG, L., INGMAN, M., SHARPE, T., KA, S., HALLBÖÖK, F., BESNIER, F., CARLBORG, Ö., BED'HOM, B., TIXIER-BOICHARD, M., JENSEN, P., SIEGEL, P., LINDBLAD-TOH, K. & ANDERSSON, L. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature,* 464**,** 587 - 593.
SIMEONE, R., MISZTAL, I., AGUILAR, I. & VITEZICA, Z. G. 2012. Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure. *Journal of Animal Breeding and Genetics,* 129**,** 3-10.
SUNDSTROM, H., WEBSTER, M. T. & ELLEGREN, H. 2004. Reduced Variation on the Chicken Z Chromosome. *Genetics,* 167**,** 377-385.
TENESA, A., KNOTT, S., WARD, D., SMITH, D., WILLIAMS, J. & VISSCHER, P. 2003. Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. *Journal of Animal Science,* 81**,** 617-623.
VAN DER WERF, J. Optimizing design of breeding programs.
VANRADEN, P. M. 2007. Genomic Measures of Relationship and Inbreeding. *Interbull Bulletin,* 37**,** 33 - 36.
VANRADEN, P. M. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science,* 91**,** 4414-4423.
VILLANUEVA, B., PONG-WONG, R., FERNÁNDEZ, J. & TORO, M. A. 2005. Benefits from marker-assisted selection under an additive polygenic genetic model. *Journal of Animal Science,* 83**,** 1747-1752.
VILLANUEVA, B., PONG-WONG, R. & WOOLLIAMS, J. A. 2002. Marker assisted selection with optimised contributions of the candidates to selection. *Genetics Selection Evolution,* 34**,** 679-704.
VISSCHER, P. M., MEDLAND, S. E., FERREIRA, M. A., MORLEY, K. I., ZHU, G., CORNES, B. K., MONTGOMERY, G. W. & MARTIN, N. G. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS genetics,* 2**,** e41.
VISSCHER, P. M., WOOLLIAMS, J. A., SMITH, D. & WILLIAMS, J. L. 2002. Estimation of Pedigree Errors in the UK Dairy Population using Microsatellite Markers and the Impact on Selection. *Journal of Dairy Science,* 85**,** 2368-2375.
WHITTAKER, J. C., THOMPSON, R. & DENHAM, M. C. 2000. Marker-assisted selection using ridge regression. *Genetics Research,* 75**,** 249-252.
WOLC, A., STRICKER, C., ARANGO, J., SETTAR, P., FULTON, E., O'SULLIVAN, N., HABIER, D., FERNANDO, R., GARRICK, D., LAMONT, S. & DEKKERS, J. C. M. Breeding Value Prediction For Production Traits In Layers Using Pedigree and Marker Based Methods. World Congress on Genetics Applied to Livestock Production, 2010.
WOLC, A., STRICKER, C., ARANGO, J., SETTAR, P., FULTON, J., O'SULLIVAN, N., PREISINGER, R., HABIER, D., FERNANDO, R., GARRICK, D., LAMONT, S. & DEKKERS, J. 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution,* 43**,** 5.

WOOLLIAMS, J. A., BIJMA, P. & VILLANUEVA, B. 1999. Expected Genetic Contributions and Their Impact on Gene Flow and Genetic Gain. *Genetics,* 153**,** 1009-1020.

WRIGHT, S. 1922. Coefficients of Inbreeding and Relationship. *The American Naturalist,* 56**,** 330-338.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. & VISSCHER, P. M. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet,* 42**,** 565-569.

CHAPTER 2

QUANTITATIVE ANALYSIS OF SIX KEY TRAITS IN BROILER PRODUCTION USING MODELS ACCOUNTING FOR MATERNAL EFFECTS

1. INTRODUCTION

Poultry products are currently the most popular of all livestock products around the world, with production levels rising from $3.8x10^6$ chickens produced worldwide in 1961 to $21.9x10^6$ produced in 2012 (FAO, 2014). This huge increase in demand for chicken products is met not only through increasing numbers of animals produced, but also by the amount of the product obtained per animal, with body weight gain per animal per day rising from 22g in the 1960s to 50g in 2000 (Arthur and Albers, 2003). This spectacular improvement was achieved partly by changes in management routines, however a large proportion of the gains was achieved through genetic improvement programme (Arthur and Albers, 2003). In practice, all highly specialized, pure lines and great-grandparent broiler stocks are owned by large, multinational breeding companies (Arthur and Albers, 2003). These stocks are subject to selection for particular production traits, on the basis of their estimated breeding value (EBV) (Wezyk and Jankowski, 2003). In meat producing chickens, the selection has usually been conducted in two stages, with the first one based on growth traits, with juvenile body weight being the leading trait, followed by selection for reproductive traits, e.g. egg weight and egg production (Wezyk and Jankowski, 2003). The response to selection and thus global production and profitability of the industry depends on the accuracy with which the EBVs are estimated.

The complexity and intensity of the selection on chickens demands highly accurate and constantly monitored statistical models for EBV estimation. Quantitative estimation of an animal's breeding value is based on its phenotypic value, determined by both its environment and breeding value, with the latter being estimated from the data on the animal and its relatives (Simm, 2000). The exact specification of the effects used in models for a given trait differs between studies and populations, as their choice is dependent on population parameters and particular selection objectives. Yet the choice of factors and effects used is crucial for unbiased estimates of trait variance components which are then used for calculation of heritability and calculation of EBVs. The estimates obtained in

particular studies can thus differ, as the genetic variance differs between breeds and even between lines of the same breed (Danbaro et al., 1995). Therefore, a careful examination of trait variance components is required for each breed and line. Furthermore, as the poultry industry is a very dynamic sector, using both intense selection and constant improvements in management of the environment, the models of variance decomposition need to be re-evaluated frequently (Koerhuis and McKay, 1996), due to changes in additive variance brought about by selection (Falconer and Mackay, 1996). Using models not fitted for the population at a specific time and consequently inadequate estimates of the variance is likely to result in diminished response to selection.

This study will aim at partitioning the variance of key broiler production traits at a scale comparable to only a few reports in the literature (Koerhuis and McKay, 1996), with the population numbers at almost 590,000 and generation numbers reaching up to 11. The selection response in chickens has been tremendous, with modern birds being three to five times heavier than their non-selected counterparts of the same age (Havenstein et al., 2003). Due to the fast progress, the number of generations used and the time since the study by Koerhuis and McKay(1996), this study provides the most current estimates of the variance components in broilers. In addition, in the presented analyses special attention will be given to dam effects on the trait, both environmental and genetic, as their proper identification could improve the estimates of the direct additive variance and thus improve the response to selection (Koerhuis et al., 1997). Numerous studies of broiler traits ignore maternal variation (e.g. Gaya et al., 2006, Le Bihan-Duval et al., 2008), thus introducing a possible bias into the estimates of variances and heritability.

2. MATERIALS AND METHODS

2.1 PHENOTYPES

The phenotypic measurements of 589,385 animals (286,838 males and 302,547 females) from a commercial broiler dam line were obtained from Aviagen Ltd. There were 30 males that had a record for a female specific trait (LFI), which were completely removed from the data set. The remaining data includes records for following traits:

- **BWT** – body weight (g) at 35 days, on both males and females, available for all individuals.

- **LFI** – feed intake (g) between 14 and 35 days, measured in feeding stations on females only, before juvenile selection.

- **AFI** – feed intake (g) between 35 and 49 days adjusted for starting weight (BWT), measured in test cages on males selected during juvenile selection at 35 days.

- **WTG** – weight gain (g) between 35 and 49 days, measured in test cages on males selected at juvenile selection.

- **EWT** – average weight (g) of eggs laid in the 48th week. The weighing is done after the eggs were stored for up to a week. The trait is recorded for females that were selected both during juvenile and adult selection, and were alive at 48 weeks.

- **HHP** – the cumulative egg production (number of eggs) during the whole laying period, recorded for each hen.

Table 2.1 gives the numbers of records available for each trait:

Table 2.1 Average trait values, with standard deviation and the number of available records per trait

Trait (unit)	Mean	Standard deviation	Available records
BWT (g)	2,008	241.8	589,385
LFI (g)	234.7	24.24	71,095
AFI (g)	184.0	22.01	25,975
WTG (g)	108.6	17.23	25,975
EWT (g)	65.11	4.10	6,345
HHP (n)	124.6	25.64	10,882

The data included also management information combined in a factor coded '**hwumgs**', which represents **hatch week**, **unit** (recorded for females only and codes for feeding station pens), **mating group** (hatch week of the parents) and **sex**. This factor had 6,002 levels, and using Python programming language has been disentangled into singular effects. The dataset contained also information on the **genotyping status** of an animal, and the **age of eggs in days** (REGEWT).

The individuals came from 339 hatch weeks, with min 693 and max 2,259 individuals in a hatch (average 1,739). The distribution of individuals over hatches is presented in Figure 2.1.

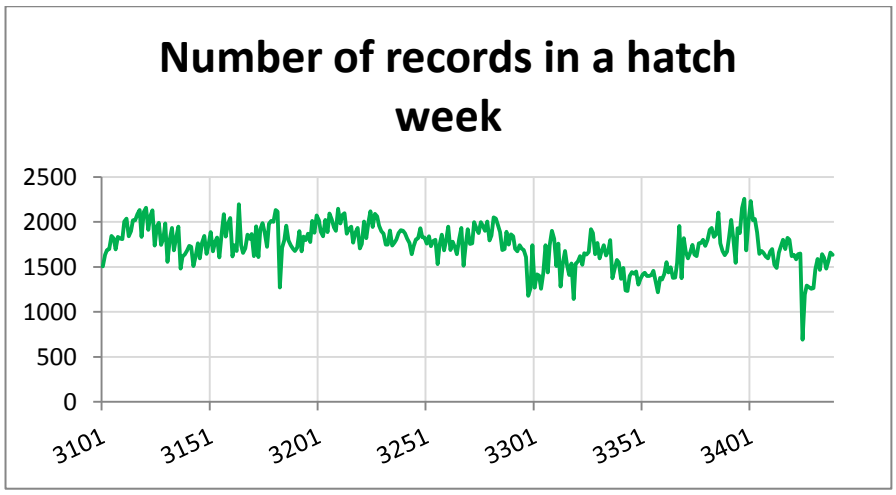

Figure 2.1 Distribution of individuals over hatch weeks

The mating group, which is also the hatch week of the parents, had 97 levels, with min 98 and max 11,902 individuals per group (average 6,017). The females were kept in 8 feeding station units, with variable number of individuals in a unit. The males were kept in pens together with their hatch week contemporaries, until the juvenile selection at 35 days, after which the selected males were placed in individual test cages.

2.2 PEDIGREE

The pedigree was constructed using the parental information provided in phenotypic records. In total the pedigree consisted of 591,199 individuals, progeny of 1,697 sires (579 sires of sires, 909 dams of sires) and 13,038 dams (944 sires of dams and 2,830 dams of dams). The average numbers of offspring were 347 (range 1 - 1,056) per sire and 45 (range 1 - 134) per dam.

The distribution of generations was continuous and overlapping, with the maximum number of generations reaching 11.

2.3 VARIANCE AND HERITABILITY ESTIMATION

2.3.1 UNIVARIATE ANALYSIS

The estimation of variance components for the broiler production traits was done by fitting mixed linear models in ASReml (Gilmour et al., 2006) which, in general form, can be described by equation 1.

Equation 1

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e}$$

Where $\boldsymbol{y}$ is the vector of observations, $\boldsymbol{\tau}$ is a vector of fixed effects, $\mathbf{X}$ is an incidence matrix referring the observations to pertaining fixed effect levels described further below, $\boldsymbol{u}$ is a vector of breeding values treated as random effects with the distribution assumptions outlined below for each model, $\boldsymbol{Z}$ is an incidence matrix referring observations to their corresponding random effects, and $\boldsymbol{e}$ is a vector of residual effects, assumed to be normally distributed with parameters $N(0, \sigma_e^2 \boldsymbol{I})$, where σ_e^2 is the residual variance and $\boldsymbol{I}$ is identity matrix.

Table 2.2 presents the initial models from which variations in relation to random effects were applied.

Table 2.2 Initial models, where *mu* is the mean of the trait in the population, *hwumgs* is the combined management effect, *p.e.* is the permanents environment effect, *hatch* is a hatch week, *egg_age* is the age of egg in days fitted as covariate and *animal* is the direct genetic effect of an individual.

Trait	Model
BWT	= mu + hwumgs + p.e. + animal
LFI	= mu + hwumgs + p.e. + animal
AFI	= mu + hwumgs + p.e. + animal
WTG	= mu + hwumgs + p.e. + animal
EWT	= mu + hatch + egg_age + animal
HHP	= mu + hatch + animal

2.3.1.1 Random Effects:

To best capture all sources of variance, several models were tested, differing in random effects used for characterizing $\boldsymbol{u}$:

- A – Direct additive genetic effect of a chick is fitted as the only random effect. In the animal model vector of random effects $\boldsymbol{u}$ is assumed to be normally distributed with parameters $N(0, \sigma_A^2 \boldsymbol{A})$, where σ_A^2 is the additive genetic variance and $\mathbf{A}$ is a numerator relationship matrix. The heritability was estimated using Equation 2.

Equation 2

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}$$

- M – Maternal genetic effect is fitted as an additional random term, next to the direct additive effect of a chick. The model used is presented in equation 3, where $\boldsymbol{Z_2}$ is an incidence matrix and $\boldsymbol{u_2}$ is vector of random dam effect, with parameters $\mathrm{N}(0, \sigma_M^2 \boldsymbol{A})$

with σ_M^2 being the dam variance. The covariance between direct and maternal effects in this model was assumed to be zero.

Equation 3

$$\boldsymbol{Y = X\tau + Z_1 u_1 + Z_2 u_2 + e}$$

The additive and maternal heritabilities were estimated as shown in Equation 4 and 5.

Equation 4

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_M^2 + \sigma_e^2}$$

Equation 5

$$m^2 = \frac{\sigma_D^2}{\sigma_A^2 + \sigma_M^2 + \sigma_e^2}$$

- ideM – The random effects include the direct genetic effect of the chick and the permanent environment of its dam. This model was used as a baseline, most often reported in the literature for variance estimation in poultry. The general form of the model can be represented in the same form as in Equation 3, but here $\boldsymbol{Z_2}$ and $\boldsymbol{u_2}$ relate to the random effects of the permanent environment, with $\boldsymbol{u_2}$ ~ N(0, $\sigma_{p.e.}^2 \boldsymbol{I}$). The heritability and proportion of the variance explained by the permanent environment were calculated as in equations 4 and 5, with the dam variance replaced by p.e. variance.

- ME – This model combined models M and ideM, with both maternal genetic and permanent environmental effects fitted along with the direct genetic effect, as represented in Equation 6, where $\boldsymbol{Z_1}$ and $\boldsymbol{u_1}$ relate to the direct additive effect of the chick with $\boldsymbol{u_1} \sim N(0, \sigma_A^2 \boldsymbol{A})$, $\boldsymbol{Z_2}$ and $\boldsymbol{u_2}$ relate to maternal additive effect of the dam, with $\boldsymbol{u_2}$ ~ N$(0, \sigma_M^2 \boldsymbol{A})$; and $\boldsymbol{Z_3}$ and $\boldsymbol{u_3}$ represent permanent environment, with $\boldsymbol{u_3}$ ~ N(0, $\sigma_{p.e.}^2 \boldsymbol{I}$). As in model M, the covariance between maternal and direct genetic effects was assumed to be zero.

Equation 6

$$\boldsymbol{Y} = \boldsymbol{X\tau} + \boldsymbol{Z_1 u_1} + \boldsymbol{Z_2 u_2} + \boldsymbol{Z_3 u_3} + e$$

- AcM – This model is equivalent to the ME model, however the covariance between maternal and direct genetic effects was to be estimated.

The heritability, maternal heritability and correlation between maternal and direct additive effects in this model were estimated as shown in Equations 7, 8, and 9 respectively.

Equation 7

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_M^2 + cov_{AM} + \sigma_{p.e.}^2 + \sigma_e^2}$$

Equation 8

$$m^2 = \frac{\sigma_D^2}{\sigma_A^2 + \sigma_M^2 + cov_{AM} + \sigma_{p.e.}^2 + \sigma_e^2}$$

Equation 9

$$r = \frac{cov_{AM}}{\sqrt{\sigma_A^2 \sigma_M^2}}$$

Using the estimates of direct and maternal additive variances and their covariance, Willham's heritability was also calculated (Willham, 1972). The calculation for this heritability is presented in Equation 10.

Equation 10

$$h_W^2 = \frac{(\sigma_A^2 + 0.5\sigma_M^2 + 1.5cov_{AM})}{\sigma_A^2 + \sigma_M^2 + cov_{AM} + \sigma_{p.e.}^2 + \sigma_e^2} = \frac{(\sigma_A^2 + 0.5\sigma_M^2 + 1.5cov_{AM})}{\sigma_P^2}$$

As the Willham's heritability can be understood as the realized heritability for mass selection, which is uncommon in modern poultry breeding, the heritable variance including maternal effect was also estimated using the method presented by Eaglen and Bijma (2009) as shown in Equation 11.

Equation 11

$$T = \frac{(\sigma_A^2 + 2cov_{AM} + \sigma_M^2)}{\sigma_P^2}$$

- AcMnope – This model was developed for traits in which maternal and direct additive effects would prove to be significant, while permanent environment would not. As such, the equations used were equivalent to equations 7 – 9, with exclusion of p.e. effects.

The analyses were run until parameters and likelihood converged, which, through a default setting in ASReml was assumed when the variance component estimates changed by no more than 1% between iterations, and the change in the likelihood was less than 0.002* current iteration number (Gilmour et al., 2006).

The significance of random terms was tested via the LogLikelihood Ratio Test (LRT), with the parameter deemed significant when twice the difference between LogL value of the model containing it and a simpler model without it exceeded 3.84.

2.3.1.2 Fixed Effects:

In addition to testing of the random effects, some options were tested in relation to fixed effects. These modifications were suited individually for particular traits.

1. The analysis of HHP and EWT was repeated with hwumgs fitted as a fixed effect instead of hatch.
2. As the feed intake is dependent on the body weight of an animal in question, the analyses of the best models for LFI were repeated, fitting BWT as a covariate.

2.3.2 BIVARIATE ANALYSIS

2.3.2.1 Male and Female BWT

The body weight was the only trait in this dataset that was recorded for both males and females. To check for the significance of the difference in variance of body weight between sexes a bivariate analysis was performed, with male and female records treated as two separate traits. In bivariate analyses $\boldsymbol{u}$ is assumed MVN(0,$\boldsymbol{V} \otimes \boldsymbol{A}$), where $\boldsymbol{V}$ is a (co)variance matrix of male and female terms.

The residual covariance between male and female body weight was set to 0, since no animal could have observations as a male and female. Equation 12 gives the formula for estimating

the genetic correlation between traits, where cov_{XY} is the genetic covariance between the two traits *X* and *Y*, and var_X and var_Y are the additive variances of the two traits.

Equation 7

$$r_g = \frac{cov_{XY}}{\sqrt{var_X var_Y}}$$

2.3.2.2 Bivariate analyses of other traits

The best fitting models identified in univariate analyses were subsequently used in bivariate analyses of pairs of traits. Table 2.3 summarizes the models used.

Due to constraints in matrix design, the bivariate analyses including EWT were run on AcM (AcMnope for EWT and HHP) model, despite other models being better for other traits. For these traits the correlations between additive and maternal variances between traits were also estimated.

Table 2.3 Choice of models fitted to traits in bivariate analyses.

	LFI	AFI	WTG	EWT	HHP
BWT	BWT ME LFI ME	BWT ME AFI ME	BWT ME WTG ME	BWT AcM EWT AcMnope	BWT ME, HHP A
LFI		LFI ME AFI ME	LFI ME WTG ME	LFI AcM EWT AcMnope	LFI ME, HHP A
AFI			AFI ME WTG ME	AFI AcM EWT AcMnope	AFI ME, HHP A
WTG				WTG AcM EWT AcMnope	WTG ME, HHP A
EWT					EWT AcMnope HHP AcMnope

Depending on models used, different correlations were estimated for particular analyses. The correlations were calculated as shown in Equation 12. The X and Y represented respectively:

- direct additive variances (σ_A^2) of the two traits when direct additive genetic correlation was calculated (r_a)
- maternal genetic variances (σ_D^2) of the two traits when maternal genetic correlation was calculated (r_m)
- permanent environment ($\sigma_{p.e.}^2$) of the two traits when correlation between permanent environment effects was calculated ($r_{p.e.}$)
- residual (σ_e^2) of the two traits when residual correlation was calculated (r_e).

The additive genetic correlation (r_a) was estimated for all pairwise comparisons, the rest of the correlations depended on the availability of the variance components (choice of model, lack of residual covariance between male and female traits etc.).

3. RESULTS

3.1 UNIVARIATE ANALYSIS

3.1.1 RANDOM TERMS

Table 2.4 shows the estimates of the heritability (h^2) obtained from models best fitting the data, and the proportion of heritable variance estimated via Willham's heritability (h_W^2) and and Eaglen and Bijma (2009) total heritability (T^2). The heritabilities obtained using the different methods showed good agreement.

Table 2.5 presents the proportion of the total variance explained by the random terms in different models. In general, the heritability estimates were found to be moderate to high, with lowest value estimated for weight gain at 0.16 (SE 0.02), and highest for EWT at 0.79 (SE 0.06). The magnitude of maternal genetic and environmental effects was similar, but when they were estimated together in the same model, the estimate of permanent environment was higher in most cases, with the exception of LFI corrected for BWT and WTG AcM, where the maternal genetic effect was slightly higher than the p.e.; and EWT and HHP, for which p.e. was estimated at 0. Except for EWT and HHP, maternal permanent environment effect was found significant for all the traits. The correlation between maternal and direct genetic effects was always negative, but significant only for EWT at -0.46 (SE 0.12).

Table 2.4 The estimates of heritability in classical sense (h^2), Willham's heritability (h_W^2) and total heritability calculated from Eaglen and Bijma (2009) (T^2).

Trait	Best Model	h^2	h_W^2	T^2
BWT	ME	0.37 (0.01)	0.38 (0.01)	0.39 (0.01)
LFI	ME	0.41 (0.02)	0.42 (0.02)	0.42 (0.02)
AFI	ME	0.23 (0.02)	0.23 (0.02)	0.23 (0.02)
WTG	ME	0.16 (0.02)	0.16 (0.02)	0.16 (0.02)
EWT	AcMnope	0.79 (0.06)	0.75 (0.05)	0.72 (0.04)
HHP	A	0.23 (0.03)	0.23 (0.03)	0.24 (0.03)

Table 2.5 Heritability estimates under different models, A - chick as the only random effect, M - chick and maternal genetic effect as random, ideM - chick and permanent environment (p.e.), ME - chick, maternal genetic and p.e., ME (BWT) – ME analysis of feed intake adjusted for BWT, AcM same effects as ME with covariance between direct and maternal genetic effects, AcMnope – direct and maternal genetic and their covariance without permanent environment. The best models are underlined. WTG AcM* - the benefit of including covariance, compared to WTG ME was significant at 0.1, but not 0.05.

Trait	Model	h^2 (SE)	m^2 (SE)	p.e. (SE)	r (SE)
BWT	A	0.54 (0.01)			
	M	0.39 (0.01)	0.05 (0.00)		
	ideM	0.39 (0.01)		0.03 (0.00)	
	ME	0.37 (0.01)	0.02 (0.00)	0.02 (0.00)	
	AcM	0.37 (0.01)	0.02 (0.00)	0.02 (0.00)	-0.02 (0.06)
LFI	A	0.53 (0.01)			
	M	0.43 (0.02)	0.04 (0.01)		
	ideM	0.43 (0.01)		0.04 (0.00)	
	ME	0.41 (0.02)	0.02 (0.01)	0.03 (0.00)	
	ME (BWT)	0.37 (0.01)	0.02 (0.00)	0.01 (0.00)	
	AcM	0.43 (0.02)	0.02 (0.01)	0.03 (0.00)	-0.16 (0.09)
AFI	A	0.27 (0.02)			
	M	0.24 (0.02)	0.02 (0.01)		
	ideM	0.24 (0.02)		0.02 (0.01)	
	ME	0.23 (0.02)	0.01 (0.01)	0.01 (0.01)	
	AcM	0.23 (0.02)	0.01 (0.01)	0.01 (0.01)	-0.18 (0.17)
WTG	A	0.21 (0.02)			
	M	0.17 (0.02)	0.02 (0.01)		
	ideM	0.16 (0.02)		0.02 (0.01)	
	ME	0.16 (0.02)	0.01 (0.01)	0.02 (0.01)	
	AcM*	0.18 (0.02)	0.02 (0.01)	0.02 (0.01)	-0.35 (0.14)
EWT	A	0.68 (0.03)			
	M	0.66 (0.03)	0.02 (0.01)		
	ideM	0.67 (0.03)		0.01 (0.01)	
	ME	0.66 (0.03)	0.02 (0.01)	0.00 (0.00)	
	AcM	0.79 (0.06)	0.04 (0.02)	0.00 (0.00)	-0.46 (0.12)
	AcMnope	0.79 (0.06)	0.04 (0.02)		-0.46 (0.12)
HHP	A	0.25 (0.02)			
	M	0.22 (0.03)	0.02 (0.01)		
	ideM	0.23 (0.02)		0.01 (0.01)	
	AcM	0.22 (0.03)	0.01 (0.01)	0.00 (0.01)	0.02 (0.26)

3.1.2 Fixed Effects

Fitting the 'hwumgs' factor in EWT resulted in 4% decrease in the additive genetic variance, with a simultaneous 7% increase in the error variance. In HHP, fitting this factor also reduced the additive variance (reduction by 9%), with minimal change in the residual variation (0.3% increase in the estimate), as shown in Table 2.6.

Table 2.6 The estimates of variance components in EWT and HHP, obtained with different fixed effects used in the model. σ_A^2 - variance of the direct genetic effect of an individual, σ_M^2 - variance of the maternal genetic effect, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance. Standard errors are given in brackets.

Trait	Model	σ_A^2 (SE)	σ_M^2 (SE)	σ_e^2 (SE)	σ_P^2 (SE)
EWT	Hwumgs	13.28 (1.55)	0.64 (0.34)	4.77 (0.78)	17.35 (0.55)
	hatch week	13.80 (1.26)	0.68 (0.26)	4.44 (0.62)	17.50 (0.48)
HHP	Hwumgs	137.70 (16.39)	-	451.54 (12.51)	598.23 (10.78)
	hatch week	151.00 (14.90)	-	450.06 (11.18)	601.06 (9.89)

3.2 Bivariate Analysis

3.2.1 Male and Female Body Weight

The analysis of body weight as separate male and female traits, with mean values of 2,170g and 1,853 respectively, confirmed that the two traits are highly correlated, albeit the correlation was lower than 1, at 0.94 (SE 0.00). The heritability estimates were also significantly different, with male heritability estimated lower at 0.34 (SE 0.01), compared to 0.41 (SE 0.01) estimated from female data.

3.2.2 Bivariate analyses

The estimates of heritability of the traits obtained from bivariate analyses did not differ largely from the ones obtained from univariate analyses, as shown in Table 2.8. The biggest difference was observed for heritability of WTG, which in univariate analyses was estimated at 0.16 (SE 0.017) while analysed together with BWT was estimated at 0.22 (SE 0.02). The estimates of the maternal genetic effect and permanent environment were even less affected by inclusion of additional traits, with most of the changes (if observed at all) noted in the estimates of standard errors.

The estimates of the genetic correlation of the direct effects between traits ranged from -0.45 (SE 0.08) between EWT and HHP, to 0.81 (SE 0.01) between BWT and LFI. The genetic correlation between maternal effects ranged between -0.63 (SE 0.29) for HHP and EWT, and 0.91 (SE 0.05) for BWT and LFI. The latter two traits shared also the highest correlation between permanent environment effects at 0.88 (SE 0.04), while the lowest correlation

between these effects was noted for BWT and AFI, at 0.05 (SE 0.15). The residual correlation ranged between -0.19 (SE 0.04) for BWT and EWT, and 0.79 (SE 0.03) for BWT and LFI. Tables 2.8 and 2.9 present the estimates of additive and maternal genetic effects (Table 2.8), and permanent environment and residual effects (Table 2.9).

Table 2.7 presents the estimates of correlations between direct and maternal genetic effects of different traits obtained from AcM models which were applied to bivariate analyses including EWT. Overall, the estimates of the correlation between maternal genetic effects of EWT with additive effects of other traits were small and insignificant. However, the correlation between maternal effects of BWT and LFI, and the additive effect of EWT was high and significant, at 0.68 (SE 0.06) for BWT and 0.65 (SE 0.10) for LFI.

Table 2.7 The estimates of correlations between direct and maternal genetic effects of particular traits.

Traits	*r*	SE
BWT A – EWT M	-0.13	0.14
BWT M – EWT A	0.68	0.06
LFI A - EWT M	-0.20	0.16
LFI M – EWT A	0.65	0.10
AFI A – EWT M	0.22	0.19
AFI M – EWT A	-0.16	0.19
WTG A – EWT M	0.06	0.22
WTG M – EWT A	0.20	0.16
HHP A – EWT M	-0.07	0.19
HHP M – EWT A	0.16	0.18

Table 2.8 The estimates of additive, h^2 (above diagonal), and maternal, m^2 (below diagonal), heritabilities with corresponding correlations (r_a and r_m respectively), obtained from bivariate analyses. Values on diagonal are the h^2 and m^2 estimates obtained in the univariate analyses with the best models fitted. "Model" rows remind which models were used.

Trait	BWT	LFI	AFI	WTG	EWT	HHP
BWT	h^2 = 0.37 (0.01) m^2 = 0.02 (0.00)	h^2 BWT = 0.37 (0.01) h^2 LFI = 0.40 (0.01) **r_a = 0.81 (0.01)**	h^2 BWT = 0.37 (0.01) h^2 AFI = 0.23 (0.02) **r_a = 0.07 (0.04)**	h^2 BWT = 0.37 (0.01) h^2 WTG = 0.22 (0.02) **r_a = 0.58 (0.04)**	h^2 BWT = 0.37 (0.01) h^2 EWT = 0.81 (0.05) **r_a = 0.32 (0.04)**	h^2 BWT=0.37 (0.01) h^2 HHP=0.24 (0.02) **r_a = - 0.22 (0.04)**
Model	ME	ME	ME	ME	BWT AcM, EWT AcMnope	BWT ME, HHP A
LFI	m^2 LFI = 0.02 (0.00) m^2 BWT = 0.02 (0.00) **r_m= 0.78 (0.05)**	h^2 = 0.41 (0.02) m^2 = 0.02 (0.01)	h^2 LFI=0.41 (0.02) h^2 AFI=0.23 (0.02) **r_a = 0.40 (0.05)**	h^2 LFI=0.41 (0.02) h^2 WTG=0.16 (0.02) **r_a=0.18 (0.06)**	h^2 LFI=0.43 (0.02) h^2 EWT=0.77 (0.05) **r_a=0.16 (0.05)**	h^2 LFI=0.41 (0.02) h^2 HHP=0.25 (0.02) **r_a = - 0.09 (0.04)**
Model	ME	ME	ME	ME	LFI AcM EWT AcMnope	LFI ME, HHP A
AFI	m^2 AFI = 0.01 (0.01) m^2 BWT = 0.02 (0.00) **r_m = - 0.22 (0.23)**	m^2 AFI = 0.01 (0.01) m^2 LFI = 0.02 (0.00) **r_m = - 0.22 (0.29)**	h^2 = 0.23 (0.02) m^2 = 0.01 (0.01)	h^2 AFI=0.22 (0.02) h^2 WTG=0.15 (0.02) **r_a=0.77 (0.00)**	h^2 AFI=0.24 (0.02) h^2 EWT=0.79 (0.06) **r_a = - 0.04 (0.09)**	h^2 AFI=0.24 (0.02) h^2 HHP=0.25 (0.00) **r_a = - 0.06 (0.07)**
Model	ME	ME	ME	ME	AFI AcM, EWT AcMnope	AFI ME, HHP A
WTG	m^2 WTG = 0.01 (0.01) m^2 BWT = 0.02 (0.00) **r_m = 0.15 (0.20)**	m^2 WTG = 0.01 (0.01) m^2 LFI = 0.02 (0.00) **r_m = - 0.13 (0.27)**	m^2 WTG = 0.01 (0.01) m^2 AFI = 0.01 (0.01) **r_m= 0.91 (0.10)**	h^2 = 0.16 (0.02) m^2 = 0.01 (0.01)	h^2 WTG=0.17 (0.02) h^2 EWT=0.79 (0.06) **r_a = - 0.06 (0.10)**	h^2 WTG=0.16 (0.02) h^2 HHP=0.26 (0.02) **r_a=-0.26 (0.07)**
Model	ME	ME	ME	ME	WTG AcM, EWT AcMnope	WTG ME, HHP A
EWT	m^2 EWT = 0.03 (0.01) m^2 BWT = 0.02 (0.00) **r_m = - 0.16 (0.19)**	m^2 EWT = 0.04 (0.01) m^2 LFI = 0.02 (0.00) **r_m = - 0.38 (0.23)**	m^2 EWT = 0.04 (0.01) m^2 AFI = 0.01 (0.01) **r_m= -0.21 (0.35)**	m^2 EWT = 0.04 (0.01) m^2 WTG = 0.02 (0.09) **r_m= 0.08 (0.32)**	h^2 = 0.79 (0.06) m^2 = 0.04 (0.02)	h^2 EWT=0.78 (0.06) h^2 HHP=0.23 (0.03) **r_a = - 0.45 (0.08)**
Model	EWT AcMnope, BWT AcM	EWT AcMnope, LFI AcM	EWT AcMnope, AFI AcM	EWT AcMnope, WTG AcM	AcM nope	AcMnope
HHP	m^2 HHP = NA m^2 BWT = 0.02 (0.00) **r_m = NA**	m^2 HHP = NA m^2 LFI = 0.02 (0.00) **r_m= NA**	m^2 HHP = NA m^2 AFI = 0.00 (0.00) **r_m= NA**	m^2 HHP = NA m^2 WTG = 0.01 (0.01) **r_m= NA**	m^2 HHP = 0.02 (0.01) m^2 EWT = 0.04 (0.01) **r_m= -0.63 (0.27)**	h^2 = 0.23 (0.03) m^2 = NA
Model	HHP A, BWT ME	HHP A, LFI ME	HHP A, AFI ME	HHP A, WTG ME	AcMnope	A

Table 2.9 Permanent environment (p.e.) estimates from bivariate analyses with their correlations, r_c (above diagonal) and residual correlations, r_e (below diagonal). Values on diagonal are the p.e. estimates obtained in the univariate analyses with the best models fitted. "Model" rows remind which models were used.

Trait	BWT	LFI	AFI	WTG	EWT	HHP
BWT	p.e. = 0.02 (0.00)	p.e. BWT =0.02 (0.00) p.e. LFI = 0.01 (0.00) **r_c= 0.88 (0.04)**	p.e. BWT = 0.02 (0.00) p.e. AFI = 0.01 (0.01) **r_c= 0.05 (0.15)**	p.e. BWT = 0.02 (0.00) p.e. WTG = 0.01 (0.01) **r_c= 0.50 (0.18)**	p.e. BWT = 0.02 (0.00) p.e. EWT = NA **r_c= NA**	p.e. BWT = 0.02 (0.002) p.e. HHP = NA **r_c= NA**
Model	ME	ME	ME	ME	BWT AcM, EWT AcMnope	BWT ME, HHP A
LFI	**r_e= 0.79 (0.03)**	p.e. =0.03 (0.00)	p.e. LFI = 0.03 (0.00) p.e. AFI = 0.01 (0.01) **r_c=0.28 (0.24)**	p.e. LFI = 0.03 (0.00) p.e. WTG = 0.02 (0.01) **r_c=0.51 (0.24)**	p.e. LFI = 0.03 (0.00) p.e. EWT = NA **r_c=NA**	p.e. LFI = 0.03 (0.00) p.e. HHP = NA **r_c= NA**
Model	ME	ME	ME	ME	LFI AcM, EWT AcMnope	LFI ME, HHP A
AFI	**r_e= -0.14 (0.01)**	**r_e= NA**	p.e. = 0.01 (0.01)	p.e. AFI = 0.01 (0.01) p.e. WTG = 0.02 (0.01) **r_c=0.77 (0.00)**	p.e. AFI = 0.02 (0.01) p.e. EWT = NA **r_c= NA**	p.e. AFI = 0.02 (0.01) p.e. HHP = NA **r_c= NA**
Model	ME	ME	ME	ME	AFI AcM, EWT AcMnope	AFI ME, HHP A
WTG	**r_e= 0.14 (0.01)**	**r_e= NA**	**r_e= 0.77 (0.01)**	p.e. =0.02 (0.01)	p.e. WTG = 0.02 (0.01) p.e. EWT = NA **r_c= NA**	p.e. WTG = 0.01 (0.01) p.e. HHP = NA **r_c= NA**
Model	ME	ME	ME	ME	WTG AcM, EWT AcMnope	WTG ME, HHP A
EWT	**r_e= -0.19 (0.04)**	**r_e=-0.09 (0.05)**	**r_e= NA**	**r_e= NA**	p.e. =NA	p.e. EWT = NA p.e. HHP = NA **r_c= NA**
Model	EWT AcMnope, BWT AcM	EWT AcMnope, LFI AcM	EWT AcMnope, AFI AcM	EWT AcMnope, WTG AcM	AcM nope	AcMnope
HHP	**r_e= -0.10 (0.01)**	**r_e= -0.05 (0.02)**	**r_e= NA**	**r_e= NA**	**r_e = - 0.03 (0.05)**	p.e. =NA
Model	HHP A, BWT ME	HHP A, LFI ME	HHP A, AFI ME	HHP A, WTG ME	HHP A, EWT AcMnope	A

4. DISCUSSION

This study presents an up-to-date summary of the genetic parameters for several key broiler traits. A previous reference study, published in Poultry Breeding and Genetics (Chambers, 1990) was a very detailed overview of reports on the genetics of meat producing chickens up to the 1990's. Following this, Koerhuis and McKay (1996) published another landmark study, based on records of over 570 thousand birds, however the analyses of this study concentrated on a limited group of egg production traits. In contrast, the presented study is based on even larger number of birds, with the records collected for traits representing three key groups of breeding goals in broiler industry: growth (BWT, WTG), feed efficiency (LFI, AFI) and reproduction (EWT, HHP).

Both previous reference studies (Chambers, 1990, Koerhuis and McKay, 1996) were published in 1990's, a decade in which numerous studies on broiler trait variance composition were carried out (e.g. Koerhuis et al., 1997, Danbaro et al., 1995, Le Bihan-Duval et al., 1998). More recent studies tend to concentrate on carcass characteristics (Grosso et al., 2010), particularly fat composition and meat eating quality, which reflect the changing market, with consumers paying more attention to the quality of the meat (Yang and Jiang, 2005). However, the new emphasis put on health, liveability and meat quality is not likely to completely replace the basic selection for underlying meat production. Lack of recent, large scale studies into growth and feed efficiency suggests that the market is relying now on estimates that are more than 10 years old. In the world of poultry breeding, based on the findings of Arthurs and Albers (2003), every year of selection can bring as much as 3% improvement. Consequently, there is a large risk that the estimates of the variance published in 1990's will not be representative of the current broiler population.

The presented study is based on one of the elite broiler lines of Aviagen Ltd., one of the leading producers of poultry meat in the world. As such, although parameters estimated here are population specific, its findings have high relevance to a large part of worldwide broiler population.

4.1 RANDOM EFFECTS

The models tested in the presented study show that while routinely used permanent environment is undoubtedly an important component of the variance for several traits, the magnitude of this effect is smaller than anticipated, and its significance is not uniform across

traits. The models used by commercial companies in their routine evaluations include the direct genetic effect of the chick and the permanent environment effect of the dam (Koerhuis and Thompson, 1997), with the latter believed to include some part of the dominance effects (Misztal and Besbes, 2000). The permanent environment has been identified as significant in numerous studies and numerous traits, sometimes at relatively high values, e.g. permanent environment of EWT was previously estimated at 0.18 (SE 0.12) (Koerhuis et al., 1997). In contrast, the results presented in this Chapter show that permanent environment is significant for growth and feed efficiency traits, albeit with a low magnitude between 0.01 and 0.03, while for the reproductive traits its significance could not be established at current record number.

The magnitude of the permanent environment effects estimated in the analysed dataset is comparable to the magnitude of maternal genetic effects, which were found to be significant for all traits but HHP. These effects are ignored in most of the commercially used and published models (e.g. Koerhuis and Thompson, 1997, Danbaro et al., 1995, Gaya et al., 2006), despite their confirmed significance (e.g. Koerhuis and Thompson, 1997) and in some instances quite large magnitude, e.g. maternal genetic effects for male and female body weight at 56 days were estimated at 0.24 (SE 0.01) and 0.21 (SE 0.01) (Mignon-Grasteau, 1999). This omission is surprising, as it has been proven that the changes in a set of traits are governed not only by the direct additive effects, but also by the inheritance of the set of characters exerting maternal effects on these traits (Lande and Kirkpatrick, 1990). The rationale behind the omission of maternal effects is based on the results of a large study on juvenile body weight in broilers, in which inclusion of the maternal genetic effects and their covariance with the direct genetic effects gave the highest likelihood, but in practical application resulted only in minor re-shuffling of the variance, observed mainly between the maternal genetic and permanent environment effects (Koerhuis and Thompson, 1997). This result was also observed here for the growth and feed intake traits when permanent environment was already included in the model, however, inclusion of maternal genetic effects provides also the opportunity to select for maternal EBVs. Maternal genetic effects are an example of associative effects, where the effect of the individual's genotype is exerted on possibly many other individuals (Bijma et al., 2007). The benefits of improved maternal EBVs could thus accumulate to significant levels, particularly in poultry, where the number of offspring per dam is far larger than in other species.

Maternal genetic effects can be considered as a generalized term covering many characteristics of the dam, which exert some effect on the offspring's phenotype. Examples

of such traits with effect on BWT include egg weight loss during incubation, egg nutrient levels, eggshell porosity and number of embryonic cells (Wilson, 1991). It has been postulated that direct selection on these traits would be preferential over the use of models containing maternal effects. For example, 7 reproductive dam traits fitted as covariates for juvenile BWT explained the difference between sire and dam based models (Koerhuis and McKay, 1996), usually attributable to maternal effects (Koerhuis et al., 1997). However, most of the egg quality traits require advanced measurement methods, as reviewed by Dunn (2011). In contrast, fitting a maternal genetic effect in the model provides a simplified yet valid method of assessing the influence of the maternal genotype on the offspring, without the added costs of recording of new traits. Furthermore, for some traits it is difficult to define precisely where the maternal effects end and the additive effects of both dam and sire start. For example, all the calcium needed for development of a hatchling is provided in the eggshell and yolk (Tuan and Ono, 1986), thus it is dam dependent. However, the actual mobilization of calcium depends on the embryo (Chien et al., 2009) and as such a combination of maternal effect and both dam's and sire's additive genes may be playing a role in the calcification of the embryo skeleton and further development of a chick. By fitting both the maternal and direct effects in the same model, it is possible to differentiate between these sources of variation.

Alongside the doubts caused by small magnitude of the maternal genetic effects, this term has been dismissed in poultry breeding also due to the belief that its influence on the phenotype of the offspring diminishes with age of the chick. While the lack of significant maternal effects in HHP would support this thesis, the significant maternal effects identified in EWT in this study cast doubts on this theory, with the maternal effect for this trait estimated at 0.04 (SE 0.02). It also stands in agreement with observations from beef cattle populations, where the maternal effects were found to initially decrease, up to 180 days of age, but stabilize thereafter (Aziz et al., 2005).

Having found significant maternal genetic effects for the majority of the traits studied, the obvious next step in the analyses was to check how these effects relate to the additive variance. The covariance between maternal and direct genetic effects varies between traits in significance, value and sign, with the effect being found significant in this dataset only for EWT. Hagger (1992) noted that this trait is subject to natural selection acting against heavy egg weight, which is also linked to lowered hatchability (Hagger et al., 1986). The early estimates on the covariance between direct and maternal genetic effects in several species were reported to be strongly negative (Robinson, 1996, Meyer, 1992). The explanations for

this phenomenon included either competition for resources between the dam and the offspring (Bijma et al., 2007), or evolutionary mechanisms maintaining the intermediate optimum of the genotypes for particular traits acting as a buffer for possible environmental changes (as reviewed by Rauw et al., 1998). In poultry, the r_{AM} estimates differ between traits and studies, with the majority reporting negative values for several traits, e.g. between -0.33 and -0.68 estimated for the fertility, hatchability and survival of layer chickens based on 2,335 hens (Hartmann et al., 2002), or between -0.12 and -0.55 for several carcass traits in broilers, based on 24,000 birds (Grosso et al., 2010). The estimates of this correlation for body weight differed between studies, from -0.54 estimated on 314,000 juvenile broilers (Koerhuis and Thompson, 1997) to 0.57 (SE 0.40) for 4,000 broilers (Pakdel et al., 2002). It appears that the relationship between the direct and maternal effects is trait, and possibly population, specific. Considering the generally low magnitude of the maternal genetic effect, detection of the r_{AM} may thus require large numbers of records.

Taking into account all the points listed above, inclusion of the maternal genetic effects in routine poultry variance evaluations could improve the estimates and create a whole new dimension in which selection could be applied.

4.2 Trait Mean Values

Although the mean values of the traits are not the subject of this study, they display the constant changes that occur in poultry production. For example, the egg weight is one of the less selected upon traits in broiler production, however, a continuous change in its mean value can be observed. In 1968, a high proportion of eggs laid by broiler pullets weighted under 58g (Morris et al., 1968), in 1995 the mean egg weights were recorded between 59.8g and 62.9g (Danbaro et al., 1995) and in the presented analysis the mean egg weight is recorded at an even higher value of 65g. The increase is probably caused by its positive correlation with juvenile body weight, which is actively selected upon in broiler chickens (Koerhuis et al., 1997, Gardiner, 1973).

4.3 Trait variance

The selection procedure in broilers is usually based on two or more stages, with the first stage, usually referred to as juvenile selection, based mainly on growth traits, while the adult selection combines growth and reproductive traits, with different emphasis in the two sexes. Because selection changes variances present in a trait (Falconer and Mackay, 1996), the procedure of estimating variance components on the final selected animals is bias prone, with additional bias likely when wrong models are fitted to the data. This bias is usually

removed in the REML procedures when all the information used for the selection is included in the model (Henderson, 1975). However, the estimates of the variance components presented in this Chapter were obtained from univariate analyses, therefore they might be suffering from the selection bias. Considering the truncation of the variance at each stage of the selection, it could be speculated that the actual additive variance component is larger than estimated on the available data, particularly for EWT and HHP.

4.3.1 BODY WEIGHT

The estimate of BWT heritability at 0.37 (SE 0.01) calculated from the large dataset of current broiler population indicates that, despite the intense selection applied, the amount of the genetic variance in this trait is still considerable. In fact, the estimate found here shows remarkable agreement with the generalized estimate of 0.4 based on a review of multiple broiler studies prior to 1990's (Chambers, 1990). There are multiple reports on this trait, using different amounts of data, collected on different lines and at variable ages, analyzed using various models. It is therefore not surprising that the heritability estimates for this trait vary greatly, between 0.1 (no SE given) in White Plymouth Rock (Danbaro et al., 1995) to 0.67 (no SE given) in an experimental broiler population (van Kaam et al., 1999). Age at recording seems to have a large effect on the estimate of the heritability of BWT, with records obtained early in life providing a higher estimate of heritability than late records (Gaya et al., 2006). Considering that most of the models used in published studies concentrate on the direct effect of the chick only, it can be speculated that these reports overestimate the genetic variance present in the studied populations. In contrast, the estimate of our study is based on the best model utilizing the significant and available sources of variation, thus providing a better estimate. Also, the number of records used in the presented analysis far exceeds the population sizes of other studies, at 590K BWT records. The previous largest study included 570K records (Koerhuis and McKay, 1996), however these numbers have not been reached in subsequent studies.

4.3.2 FEED INTAKE

The difference between the estimates of heritability for female (LFI) and male (AFI) feed intake, at 0.41 (SE 0.02) and 0.23 (SE 0.02) respectively, can have several possible explanations: the true difference in feed intake between sexes, differences brought about by the correction of AFI for body weight, different period of recording, and mode of measurement with females recorded in feeding stations while roaming free with their contemporaries, vs. males maintained in individual test cages of limited space. The latter factor can be further split into several behavioural aspects, from the possible links between

feed intake and activity levels, to associative behavior of feeding. The moderate genetic correlation between LFI and AFI, estimated at 0.40 (SE 0.05), indicates that these two traits are considerably different. In practice, feed efficiency in broiler chickens is usually expressed in the form of Feed Conversion Ratio (FCR) or Residual Feed Consumption (RFC) (Szwaczkowski, 2003), calculated from the feed intake and body weight gain. While this ratio provides useful information to the breeders, its use in genetic evaluations is somewhat surprising, as the heritability of such ratios has been found to be lower than heritability of the traits considered separately (Gaya et al., 2006). The exact specification of the measurements, i.e. age and time period over which the records are collected, housing system and sex, introduce a large variability into the trait definition. As such, it is difficult to compare the results observed in this analysis to previously reported estimates. Example estimates of heritability for feed intake range between 0.2 (SE 0.03), found for a broiler sire line with records collected between 35 and 49 days of age (Gaya et al., 2006), 0.25 (no SE given) found in an experimental broiler line, measured on both sexes between 22 and 48 days of age (van Kaam et al., 1999) and 0.47 (SE 0.05) calculated based on records collected between 17 and 23 days of age, on two lines divergently selected for digestive efficiency (de Verdal et al., 2011).

Considering the logical and biological links between feed intake and the body weight, some studies suggested fitting the latter as a covariate in feed intake models (van Kaam et al., 1999). Not surprisingly, such correction brings a significant change in the estimates of the variance components, followed by changes in the heritability estimates. In the presented study such correction reduced the heritability from 0.41 (SE 0.02) to 0.37 (SE 0.01). However, the validity of such corrections is questionable when one considers the significant genetic correlation between those two traits, estimated at 0.81 (SE 0.01). In such cases, fitting the traits together in bi- or multi-variate analyses appears to be a much better option, as it fully utilizes the information on the (co)variances of included traits.

4.3.3 WEIGHT GAIN

Out of the traits analyzed in the presented study, WGT was found to have the lowest heritability, at 0.16 (SE 0.02). Van Kaam et al. (1999) reported the heritability of inferred weight gain between 23 and 48 days of age separate for males and females, at 0.23 and 0.19 respectively (no SE given), obtained from a bivariate analysis accounting for the maternal genetic effects. The review of broiler studies published between 1958 and 1986, presented by Chambers (1990), gives much higher values of heritability of weight gain between 0.42 and 0.72. It could be speculated that the reduction in the additive variance in the presented results

as compared to the published studies is an effect of the artificial selection, with weight gain being one of the key breeding goals. The improvement in this trait over the last decades has been tremendous, with an estimated 84g/year increase brought about by both the genetic and management changes (Havenstein et al., 2003). However, considering that the reduction in the additive variance of the related key broiler trait, i.e. body weight, is not as marked, it is more likely that the differences between the estimates presented in this Chapter and in studies reviewed by Chamber (1990) have other sources. Inflation of the previously published estimates due to use of models that do not account for maternal effects is one possibility, supported by a relatively good agreement of the presented results with estimates from Van Kaam et al (1999) who used a model equivalent to model M in this study. Alternatively, the differences may be simply caused by different measurement characteristics, or simply differences between lines.

4.3.4 Egg Weight and Hen Housed Production

Egg weight and hen housed production both lacked the significant influence of the permanent environment, however the best models fitted to these two traits differed considerably, with a simple model showing best fit to HHP data, and the comprehensive AcM model fitting best the EWT. Lack of significance of the permanent environment effects in reproductive traits has been identified before (e.g. Koerhuis et al., 1997, Koerhuis and McKay, 1996).

Egg weight is the only trait of those analysed in this Chapter for which inclusion of maternal effects did not introduce a marked change in the estimate of the direct genetic variance. However, fitting their covariance actually increased the estimate of heritability, from 0.66 (SE 0.03) calculated using the ME model, to 0.79 (SE 0.06) calculated from AcMnope model. This highlights the importance of testing not only for the presence of maternal effects, but also for their covariance with the direct effects.

The relative agreement between the estimates of heritability for EWT presented here and in previously published reports, e.g. 0.55 (SE 0.14) calculated using a model equivalent to ideM in study by Koerhuis et al. (1997) is consistent with low selection pressure applied to this trait, with the changes of the genetic variance for this trait over time, and between broiler lines, being negligible.

The estimates of the heritability of HHP are usually moderate (as reviewed by Chambers, 1990), however depending on the age at which they are collected, they can range between 0.07 (SE 0.01) (Farzin et al., 2013) and 0.54 (no SE given) (Luo et al., 2007). HHP is a

complex trait, influenced by several other traits, like the age at first egg, rate of egg production and hen's viability (Fairfull, 1990). The distribution of HHP records also departs from normality, with the highest likelihood of the models fitted to this trait obtained when some transformations are applied to the raw data (Koerhuis, 1996). The use of raw data in this study was aimed at mirroring the routine commercial evaluations, thus, it is possible that the genetic variance estimates found in the presented analyses would improve after transformations, as was found by Koerhuis (1996).

Further, the estimates of the genetic variance for EWT and HHP can be expected to be biased downwards, as mentioned before. These two traits are collected on adult birds, which passed several stages of selection. As such, the distribution of records collected on selected birds is not likely to represent the true, underlying distribution.

4.4 CORRELATION BETWEEN TRAITS

Significant correlations between the traits analyzed in this Chapter underline the need for multivariate models in broiler evaluations. The high correlation between BWT and LFI at 0.81 (SE 0.01), contrasted with the low correlation between BWT and AFI estimated at 0.07 (SE 0.04) indicate that the body weight at the end of a feed intake recording period is more informative than the weight recorded at its beginning. Interestingly, the estimates of the heritability for most traits did not change with the bivariate models, with the exception of WTG, for which a bivariate model with BWT increased the estimate of heritability from 0.16 (SE 0.02) calculated in univariate analyses to 0.22 (SE 0.02). However, the benefits of the bi- and multi-variate models are based mostly on improved EBV prediction (Mrode, 2005). Considering that selection introduces linkage disequilibrium which affects variances and covariances of all correlated traits (Villanueva and Kennedy, 1990), including as many traits as computationally feasible in the routine evaluations should be continued.

The positive and strong correlation between the direct effect of EWT and the maternal effect of LFI found in this study extends the previously found link between the egg weight and BWT. Considering the high correlation between BWT and LFI it seems likely that these two traits are partly determined by the same genes. Previously, the effect of the egg weight on the body weight was considered as a maternal effect (as reviewed by Koerhuis et al., 1997). However, the positive correlation between the direct effects of BWT and EWT, and the maintained presence of maternal effects for both traits even when they are fitted in a bivariate analysis, indicates that the network of direct and maternal effects between BWT, LFI and EWT is more complicated than previously expected.

5. CONCLUSIONS

The heritability of the key broiler traits analyzed in this Chapter range between 0.16 (SE 0.02) for WTG and 0.79 (SE 0.06) for EWT. The likelihood of the models indicates that permanent environment significantly affects growth and feed efficiency traits, but has very minor influence over reproductive traits. Maternal genetic effects affected both growth and reproduction traits, with the one exception of HHP. They were found in EWT, a trait recorded late in life, which indicates that maternal effects can influence performance of offspring at all life stages. Inclusion of a significant covariance between direct and maternal genetic effects significantly affects the fit of the models, and heritability estimates, as shown in EWT. Complex networks of correlations between direct and maternal effects of different traits underlines the need for the use of multivariate models.

6. REFERENCES

ARTHUR, J. A. & ALBERS, G. A. 2003. Industrial Perspective on Problems and Issues Associated with Poultry Breeding. *In:* MUIR, W. M. & AGGREY, S. E. (eds.) *Poultry Genetics, Breeding and Biotechnology.* CABI Publishing.

AZIZ, M. A., NISHIDA, S., SUZUKI, K. & NISHIDA, A. 2005. Estimation of direct and maternal genetic and permanent environmental effects for weights from birth to 356 days of age in a herd of Japanese Black cattle using random regression. *Journal of Animal Science,* 83**,** 519-530.

BIJMA, P., MUIR, W. M. & VAN ARENDONK, J. A. 2007. Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics,* 175**,** 277-88.

CHAMBERS, J. 1990. Genetics of Growth and Meat Production in Chickens. *In:* CRAWFORD, R. (ed.) *Poultry Breeding and Genetics.* Elsevier.

CHIEN, Y. C., HINCKE, M. T. & MCKEE, M. D. 2009. Ultrastructure of avian eggshell during resorption following egg fertilization. *Journal of Structural Biology,* 168**,** 527-538.

DANBARO, G., OYAMA, K., MUKAI, F., TSUJI, S., TATEISHI, T. & MAE, M. 1995. Heritabilities and Genetic Correlations from a Selection Experiment in Broiler Breeders using Restricted Maximum Likelihood. *Japanese Poultry Science,* 32**,** 257 - 266.

DE VERDAL, H., NARCY, A., BASTIANELLI, D., CHAPUIS, H., MEME, N., URVOIX, S., LE BIHAN-DUVAL, E. & MIGNON-GRASTEAU, S. 2011. Improving the efficiency of feed utilization in poultry by selection. 2. Genetic parameters of excretion traits and correlations with anatomy of the gastro-intestinal tract and digestive efficiency. *BMC Genetics,* 12**,** 71.

DUNN, I. 2011. Poultry breeding for egg quality: traditional and modern genetic approaches. *Improving the safety and quality of eggs and egg products. Volume 1: Egg chemistry, production and consumption* pp 245-260.

FAIRFULL, R. W., GOWE, R. S 1990. Genetics of egg production in chickens. *Poultry Breeding and Genetics.* Amsterdam: Elsevier.

FALCONER, D. S. & MACKAY, T. F. C. 1996. *Introduction to quantitative genetics / D.S. Falconer and Trudy F.C. Mackay*, Harlow : Longman, 1996.4th ed.

FAO, F. A. A. O. O. T. U. N. 2014. *FAO Food Statistics* [Online]. Available: http://faostat3.fao.org/faostat-gateway/go/to/download/FB/FBS/E.

FARZIN, N., TORSHIZI, R. V., GERAMI, A. & SERAJ, A. 2013. Estimates of genetic parameters for monthly egg production in a commercial female broiler line using random regression models. *Livestock Science,* 153**,** 33-38.

GARDINER, E. E. 1973. Effects of Egg Weight on Posthatching Growth Rate of Broiler Chicks. *Canadian Journal of Animal Science,* 53**,** 665-668.

GAYA, L., FERRAZ, J., REZENDE, F., MOURAO, G., MATTOS, E., ELER, J. & MICHELAN FILHO, T. 2006. Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. *Poultry Science,* 85**,** 837-843.

GILMOUR, A. R., GOGEL, B. J., CULLIS, B. R., WELHAM, S. J. & THOMPSON, R. 2006. *{ASREML} User Guide Release 2.0*.

GROSSO, J. L. B. M., BALIEIRO, J. C. C., ELER, J. P., FERRAZ, J. B. S., MATTOS, E. C. & MICHELAN, T. 2010. Comparison of different models to estimate genetic parameters for carcass traits in a commercial broiler line. *Genetics and Molecular Research,* 9**,** 908-918.

HAGGER, C., STEIGER-STAFL, D. & MARGUERAT, C. 1986. Embryonic Mortality in Chicken Eggs as Influenced by Egg Weight and Inbreeding. *Poult Sci,* 65**,** 812-814.

HARTMANN, C., STRANDBERG, E., RYDHMER, L. & JOHANSSON, K. 2002. Genetic Relations between Reproduction, Chick Weight and Maternal Egg Composition in a White Leghorn Line. *Acta Agriculturae Scandinavica, A,* 52**,** 91-101.

HAVENSTEIN, G., FERKET, P. & QURESHI, M. 2003. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poult Sci,* 82**,** 1500-1508.

HENDERSON, C. R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,* 31**,** 423-447.

KOERHUIS, A. & THOMPSON, R. 1997. Models to estimate maternal effects for juvenile body weight in broiler chickens. *Genetics Selection Evolution,* 29**,** 225 - 249.

KOERHUIS, A. N. M. 1996. Non-normality of egg production distributions in poultry and the effects of outlier elimination and transformation on size and curvilinearity of heritability. *Livestock Production Science,* 45**,** 69-85.

KOERHUIS, A. N. M. & MCKAY, J. C. 1996. Restricted maximum likelihood estimation of genetic parameters for egg production traits in relation to juvenile body weight in broiler chickens. *Livestock Production Science,* 46**,** 117-127.

KOERHUIS, A. N. M., MCKAY, J. C., HILL, W. G. & THOMPSON, R. 1997. A genetic analysis of egg quality traits and their maternal influence on offspring-parental regressions of juvenile body weight performance in broiler chickens. *Livestock Production Science,* 49**,** 203-215.

LANDE, R. & KIRKPATRICK, M. 1990. Selection Response in Traits with Maternal Inheritance. *Genetical Research,* 55**,** 189-197.

LE BIHAN-DUVAL, E., DEBUT, M., BERRI, C., SELLIER, N., SANTE-LHOUTELLIER, V., JEGO, Y. & BEAUMONT, C. 2008. Chicken meat quality: genetic variability and relationship with growth and muscle characteristics. *BMC Genetics,* 9**,** 53.

LE BIHAN-DUVAL, E., MIGNON-GRASTEAU, S., MILLET, N. & BEAUMONT, C. 1998. Genetic analysis of a selection experiment on increased body weight and breast muscle weight as well as on limited abdominal fat weight. *British Poultry Science,* 39**,** 346-353.

LUO, P. T., YANG, R. Q. & YANG, N. 2007. Estimation of genetic parameters for cumulative egg numbers in a broiler dam line by using a random regression model. *Poultry Science,* 86**,** 30-36.

MEYER, K. 1992. Variance-Components Due to Direct and Maternal Effects for Growth Traits of Australian Beef-Cattle. *Livestock Production Science,* 31**,** 179-204.

MIGNON-GRASTEAU, S. 1999. Genetic parameters of growth curve parameters in male and female chickens. *British Poultry Science,* 40**,** 44 - 51.

MISZTAL, I. & BESBES, B. 2000. Estimates of parental-dominance and full-sib permanent environment variances in laying hens. *Animal Science,* 71**,** 421-426.

MORRIS, R. H., HESSELS, D. F. & BISHOP, R. J. 1968. The relationship between hatching egg weight and subsequent performance of broiler chickens. *British Poultry Science,* 9**,** 305-315.

MRODE, R. A. 2005. *Linear models for the prediction of animal breeding values,* Wallingford: CABI

PAKDEL, A., VAN ARENDONK, J., VEREIJKEN, A. & BOVENHUIS, H. 2002. Direct and maternal genetic effects for ascites-related traits in broilers. *Poultry Science,* 81**,** 1273-1279.

RAUW, W. M., KANIS, E., NOORDHUIZEN-STASSEN, E. N. & GROMMERS, F. J. 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science,* 56**,** 15-33.

ROBINSON, D. L. 1996. Models which might explain negative correlations between direct and maternal genetic effects. *Livestock Production Science,* 45**,** 111-122.

SIMM, G. 2000. *Genetic Improvement of Cattle and Sheep,* Farming Press.

SZWACZKOWSKI, T. 2003. Use of mixed model methodology in poultry breeding: estimation of genetic parameters. *In:* MUIR, W. M. & AGGREY, S. E. (eds.) *Poultry Genetics, Breeding and Biotechnology.* CABI Publishing.

TUAN, R. S. & ONO, T. 1986. Regulation of Extraembryonic Calcium Mobilization by the Developing Chick-Embryo. *Journal of Embryology and Experimental Morphology,* 97**,** 63-74.

VAN KAAM, J., GROENEN, M., BOVENHUIS, H., VEENENDAAL, A., VEREIJKEN, A. & VAN ARENDONK, J. 1999. Whole genome scan in chickens for quantitative trait loci affecting growth and feed efficiency. *Poultry Science,* 78**,** 15-23.

VILLANUEVA, B. & KENNEDY, B. W. 1990. Effect of Selection on Genetic-Parameters of Correlated Traits. *Theoretical and Applied Genetics,* 80**,** 746-752.

WEZYK, S. & JANKOWSKI, J. 2003. Application of Mixed Model Methodology in Breeding Strategies for Meat-type Birds. *In:* MUIR, W. M. & AGGREY, S. E. (eds.) *Poultry Genetics, Breeding and Biotechnology.* CABI Publishing.

WILLHAM, R. L. 1972. The Role of Maternal Effects in Animal Breeding: III. Biometrical Aspects of Maternal Effects in Animals. *Journal of Animal Science,* 35**,** 1288-1293.

WILSON, H. R. 1991. Interrelationships of egg size, chick size, posthatching growth and hatchability. *World's Poultry Science Journal,* 47**,** 5-20.

YANG, N. & JIANG, R.-S. 2005. Recent advances in breeding for quality chickens. *World's Poultry Science Journal,* 61**,** 373-381.

CHAPTER 3

THE CHANGES IN THE GENETIC VARIANCE OF BROILER TRAITS AND THEIR EFFECT ON THE EBV PREDICTION

1. INTRODUCTION

Since domestication, livestock species have been subject to artificial selection where the individuals exhibiting the most desirable characteristics have been consciously chosen by breeders to produce the next generation. The methods of identifying the best individuals have been continuously developing as knowledge of the modes of inheritance and mathematical abilities of breeders progressed. It could be argued that this selection facilitated the development of human societies, starting from selection for tameness, which allowed the early humans to turn from hunters/gatherers to more agricultural societies. Nowadays methods of identifying the best individuals and range of selective goals are much more advanced than in those early times, however the basic definition of the process remains the same. It is perhaps best summarized by the words of Charles Darwin:

"Variation is a feature of natural populations and every population produces more progeny than its environment can manage. The consequences of this overproduction is that those individuals with the best genetic fitness for the environment will produce offspring that can more successfully compete in that environment. Thus the subsequent generation will have a higher representation of these offspring and the population will have evolved." (Darwin, 1859)

Darwin's theory of the genetic changes as a result of selection, be it natural or artificial, has found confirmation in modern studies, originally based on advanced theoretical calculations (e.g. Henderson et al., 1959) and recently using direct information on the genes (e.g. a review by Cutter and Payseur, 2013). The changes in the genetic constitution as a result of selection pressure can be explained through several mechanisms.

Firstly, the most straightforward effect of the selection is expressed through the change in allele frequencies. Selecting individuals that exhibit the most advantageous phenotypes increases the frequency of alleles that these individuals carry (Falconer and Mackay, 1996). The extent of these changes depends on the selection intensity, i.e. what proportion of the population will contribute to the next generation, initial allele frequencies and dominance at a given locus, especially when selection is based on the phenotype (Falconer and Mackay,

1996). This directional effect is frequently used to visualize the changes in the genotypes, however, under the infinitesimal model selection pressure acting on each of the infinite number of loci is extremely small. As such, changes of the mean value of the trait in the progeny of selected parents are brought through infinitesimal changes at each of the infinite number of loci, thus having a negligible effect on their variance (Bulmer, 1980).

Secondly, selection introduces linkage disequilibrium (LD) between loci of interest, thus affecting the genetic variance. A reduction of the additive genetic variance caused by negative covariances between loci, considered within the infinitesimal model, was first described by Bulmer (1971) and has been known since as "Bulmer effect". The magnitude of this effect depends on the number of loci affecting the trait and the selection intensity. In contrast to the directional gene frequency changes which are permanent even after the selection has ceased, the Bulmer effect diminishes with each non-selected generation, as recombination breaks the associations between loci, reverting the covariance between them back to zero (Bulmer, 1971).

Finally, the selection limits the numbers of the parents contributing to the next generation, narrowing already the finite size of the population. This leads to genetic drift acting across the genome, which leads to fixation/loss of some alleles at a rate depending on the population size (Falconer and Mackay, 1996). Under the infinitesimal model the direction of this process is random, as there is only a negligible selection pressure acting on any particular locus, however, the magnitude of the reduction in the genetic variance among the selected offspring is approximated as $\left(1 - {}^{1}/_{2N_e}\right)$, where N_e is the effective size of the population (Falconer and Mackay, 1996). In practice, drift is frequently considered in terms of inbreeding levels. The inbreeding coefficient (F) is defined as the probability that two alleles taken at random from the parental population are identical by descent (Falconer and Mackay, 1996). The reduction of variance in progeny due to inbreeding is estimated as (1-F) (Falconer and Mackay, 1996). Although drift and inbreeding relate to the same phenomenon of the fixation and loss of alleles, the effects of inbreeding may be intensified through the inherited selective advantage of the offspring of selected parents (Woolliams et al., 1993). The changes in the genetic variance caused by drift and inbreeding are permanent, and the proportion of the variance lost due to fixation/loss is irreversible, unless other forces, such as outcrossing or mutation, introduce new variation to the population.

The factors outlined above lead to changes in the mean and variance of a trait under selection pressure, which can be termed as progress. The rate of this progress, or gain, depends on the

intensity of selection, and accuracy with which breeders are able to estimate the genetic value of the selection candidates. This genetic value is usually referred to as the estimated breeding value (EBV).

Nowadays, most of the methods used for the prediction of the EBVs are based on Best Linear Unbiased Prediction (BLUP), a method first proposed by Henderson (1950), which calculates the EBVs based not only on the base of progeny performance, but using the information sourced from all relatives of the selection candidate. Information from the relatives is obtained by comparing the resemblance between their phenotypes and their genetic "similarity", interpreted as the numerator relationships derived from the pedigree. The covariance between the breeding values of relatives is thus obtained by scaling the matrix of covariances between relatives, quantified by the relationship coefficients, by the estimate of the genetic variance present for the trait in the given population (Henderson, 1975). The relationship coefficients are calculated from the pedigree relationships, while the estimate of the genetic variance is usually obtained by Restricted Maximum Likelihood (REML), as reviewed by Hofer (1998).

The estimate of the genetic variance (σ_A^2) approximates the amount of the genetic variance present in the base population, i.e. the population from which the studied population originated (van der Werf and de Boer, 1990, Henderson, 1985). This population is assumed to conform to the characteristics of an idealized population, i.e. it consists of individuals that are unrelated, non-inbred, originate from random mating and are unselected (Henderson, 1985). In practice, finding such a population is impossible, but the earliest generation recorded in the pedigree (i.e. individuals with unknown parents) is assumed to conform to these characteristics. Thus, by analyzing the phenotypic records collected over several generations, the estimates of the variance approximate the genetic variability present among the first generation recorded in the pedigree.

As BLUP methodology was developed mostly for livestock populations, the effect of selection on the parameters approximating the genetic variance was of high importance. Examination based on theoretical derivation showed that when all the criteria used for selection are incorporated into the analysed dataset, BLUP accounts for the reduction in variance due to the Bulmer effect (Henderson, 1975), while the effects of inbreeding are accounted for by fitting the pedigree. Because BLUP is based on an infinitesimal model, the effect of the changes in allele frequencies resulting from selection is considered to be negligible and thus is not accounted for.

In most livestock species, the selection is based on multiple traits, with different weights given to particular characteristics. For example in broiler chickens, the selection is carried out in at least two stages, where the juvenile selection is based around growth and feed efficiency traits, while the adult selection concentrates on reproductive performance. Juvenile body weight (BWT) has been the key trait in the broiler production, due to its large impact on feed conversion ratio (FCR) and consequently profits, its high heritability and relative ease of recording, where the phenotypes can be collected early in life (Arthur and Albers, 2003). As such selection on this trait could be seen as dominating the selective decisions in broilers. However, the direct effects of selection on other, possibly correlated, traits and simultaneous natural selection of the breeder birds should also be considered.

This analysis presents a thorough examination of the REML estimates of variance components and predictive ability of BLUP for two key broiler traits, BWT and hen housed production (HHP), based on a vast dataset containing 1.3M birds, spread over 24 generations. An analysis of the full dataset using models containing various random effects provides an insight into identifiable sources of variation. Estimating the genetic variance in the base population of the large pedigree, but based on the records of individuals hatched at different points in time reveals the trends in the genetic variance. Finally, the predictive ability of BLUP using such a large dataset is assessed by bias and accuracy measures obtained from a cross-validation.

2. MATERIALS

2.1 PEDIGREE

The pedigree used in the first part of the analyses (ped_{FULL}) contained 1,305,377 individuals spread over 24 generations relating to approximately 15 years of recording. There were 3,896 sires (offspring to 1,468 paternal grandsires and 2,233 paternal granddams) and 29,973 dams (offspring to 2,288 maternal grandsires and 6,540 maternal granddams), with average number of offspring at 335 (range 1 - 1,056) per sire and 43 (range 1 - 134) per dam. The base population, i.e. the earliest recorded individuals with parents assumed unknown, consisted of 225 sires and 1,653 dams. The average inbreeding coefficient for the population described by ped_{FULL} was estimated at 0.059.

To evaluate the influence of the pedigree depth, the available dataset was split into periods. The full pedigree (ped_{FULL}) was cut into two parts. The first part (ped_1) contained 643,455

individuals spread over 12 generations and hatched in the first 7 years of the recording period. Ped_1 contained 2,042 sires (offspring to 767 paternal grandsires and 1,135 paternal granddams) and 15,242 dams (offspring to 1,136 maternal grandsires and 3,048 maternal granddams), with average number of offspring at 314 (range 1 - 932) per sire and 42 (range 1 - 124) per dam. The base population for this pedigree was the same as for ped_{FULL}. The individuals contained in ped_1 were characterized by average inbreeding of 0.025.

The second part of the pedigree (ped_2) contained 663,507 individuals spread over 13 generations, which hatched in the latter 7 years of the recording period. There were 1,901 sires (offspring to 652 paternal grandsires and 1,019 paternal granddams) and 15,153 dams (offspring to 1,085 maternal grandsires and 3,271 maternal granddams), with average number of offspring at 348 (range 1 - 1,056) and 44 (range 1 - 134) per sire and per dam respectively. The base population of this pedigree consisted of 1,585 individuals, 176 sires and 1,409 dams. The average inbreeding for this group calculated from ped_{FULL} was estimated at 0.091. The average inbreeding for individuals contained in ped_2 only was estimated at 0.038.

2.2 PHENOTYPES

Two key traits in broiler production were selected for analyses: BWT - body weight at 35 days of age, measured on both males and females and HHP - hen housed production, recorded as the cumulative egg production per hen, during the whole laying period. With exception of the base generation individuals, all animals had a record for BWT, but only a fraction had records for HHP - these were females that passed several stages of selection. Along with the split of the pedigree outlined above, the phenotypic records were split into smaller parts as well. Dataset DAT_{FULL} contained all the records available, DAT_1 contained records for individuals listed in ped_1 and DAT_2 contained phenotypes of individuals listed in ped_2. Table 3.1 presents the numbers of phenotypic records for both traits in the three scenarios, together with mean and standard deviation for the included records.

Table 3.1 Mean, standard deviation and number of records for BWT and HHP in the whole data set (DAT_{FULL}) and subsets (DAT_1 and DAT_2).

Trait	DAT_{FULL}			DAT_1			DAT_2		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
BWT (g)	1,950	253.8	1,303,499	1,902	255.9	641,577	1,997	242.8	661,922
HHP	122.8	26.55	28,856	121.9	25.23	14,868	123.6	27.85	13,988

3. METHODS

Estimation of variance components, heritabilities and prediction of EBVs were carried out in ASReml 3.0 (Gilmour et al., 2006) by fitting mixed linear models to the data. The description of the methodology was presented in Chapter 2.

3.1 PEDIGREE DEPTH AND NUMBER OF RECORDS

3.1.1 MODEL CHOICE ON FULL DATA

First, several models were fitted to the DAT_{FULL} (with corresponding ped_{FULL}) to test if increasing the depth of data, as compared to results presented in Chapter 2, changes the significance of random effects. The models used were as outlined in Chapter 2. Table 3.2 recaps the list of models, with corresponding estimated parameters. While the choice of random effects differed between models, the fixed effects used in all analyses were constant for the trait, with 'hatch week' fitted to HHP, and a combined factor 'hwumgs', including hatch week, unit, mating group and sex, fitted to BWT.

Table 3.2 Models tested, with list of random effects used and parameters estimated. σ_A^2 - variance of the direct genetic effect of an individual, $\sigma_{p.e.}^2$ - variance of the permanent environment of the dam, σ_M^2 - variance of the maternal genetic effect, σ_e^2 - residual variance, $\sigma_{A,M}$ - covariance between direct and maternal genetic effects, h^2- direct heritability, *p.e.* - proportion of the total variance explained by permanent environment, m^2 - maternal heritability, r_{AM} - correlation between direct and maternal genetic effects.

Model	Variance Component Used	Parameters
A	$\sigma_A^2 + \sigma_e^2$	h^2
ideM	$\sigma_A^2 + \sigma_{p.e.}^2 + \sigma_e^2$	h^2, *p.e.*
M	$\sigma_A^2 + \sigma_M^2 + \sigma_e^2$	h^2, m^2
ME	$\sigma_A^2 + \sigma_M^2 + \sigma_{p.e.}^2 + \sigma_e^2$	h^2, m^2, *p.e.*
AcM	$\sigma_A^2 + \sigma_M^2 + \sigma_{A,M} + \sigma_{p.e.}^2 + \sigma_e^2$	h^2, m^2, r_{AM}, *p.e.*

3.1.2 CHANGES OF VARIANCE OVER TIME

To estimate how phenotypic records collected over time estimate the variance in the base population, the data was split into eight fragments, using two approaches. In the first one the split of data corresponded to the split in the pedigree, with phenotypic records placed in DAT_1 and DAT_2, as described in Materials.

The analyses where then carried out as presented in Table 3.3, ALL - included all available information (ped_{FULL} and DAT_{FULL}), P1 - included only information for earlier hatches (ped_1, DAT_1), P2 - used only information from the latter period (ped_2, DAT_2) and P2_ALL

contained phenotypes in the latter period, referred to the full pedigree (ped_{FULL} and DAT_2). The choice of random effects used for these analyses differed between traits and was based on best models identified in Chapter 1, i.e. ME for BWT and A for HHP.

Table 3.3 Pedigree and phenotypes used in the analyses

	ALL	P1	P2	P2_ALL
Pedigree	ped_{FULL}	ped_1	ped_2	ped_{FULL}
Data	DAT_{FULL}	DAT_1	DAT_2	DAT_2

The traits were initially examined separately, in univariate analyses. Afterwards, to approximate a real-life situation where several traits of interest are analyzed simultaneously and to account for possible genetic covariances between traits, a bivariate model was fitted to BWT and HHP records.

To fully utilize the number of records available, the phenotypic records were further split into 8 periods, each containing phenotypes from 2 years of recording (DAT_{Y1} - DAT_{Y8}). Those fragments were then analyzed with the ped_{FULL} by fitting ME model to BWT and A model to HHP.

3.2 Bias and Accuracy of EBV Predictions

To test the predictive ability of the models fitted to the large dataset available, a cross-validation approach was used, where EBVs for selection candidates where predicted while their actual records were masked. As the last generation contained juvenile animal records only, selection candidates were chosen as the individuals in the generation next to the last, to ensure availability of HHP records which are collected on adult birds only. The last generation animals were then removed from the analysis. Thus, the pedigree used for predictions (ped_{PRED}) contained 1,256,224 individuals spread over 23 generations. There were 67,487 selection candidates, each of which had a masked phenotype for BWT and 1,763 had records for HHP. The predictions were based on 1,183,729 BWT records and 26,795 HHP records, contained in DAT_{PRED}.

Models fitted included A, M, ideM and ME in univariate analyses. In bivariate analysis, the model fitted to HHP was a simple animal model (A), while for BWT the model fitted included also maternal effects (ME). The combined effect ‘hwumgs’ was fitted as a fixed effect to BWT and hatch week was fitted to HHP in both uni- and bi-variate analyses. For bivariate analyses, two scenarios were tested. In the first scenario, the predicted EBVs for the two traits were extracted and then separately regressed against recorded trait values

(Scenario 1). In Scenario 2, the BWT records for the selection candidates were retained in the analyses and used for the prediction of HHP values. The predictive ability in this scenario was assessed for predicted HHP EBVs only.

The bias of the predictions was calculated as a regression coefficient of the actual phenotypes recorded for the selection individuals, regressed on the predicted EBVs, with fixed effects accounted for in the model, as shown in Equation 1, where y_{ij} is the phenotype of individual j in the i^{th} fixed effect group, μ is the mean of the trait, α_i is the intercept in the i^{th} fixed effect group, β is the slope of the regression line, x_{ij} is the EBV obtained from analyses using the given model and e_{ij} is the error term for this individual.

$$y_{ij} = \mu + \alpha_i + \beta x_{ij} + e_{ij}$$

For models accounting for maternal genetic effects, the bias of maternal EBV prediction was also calculated. In such cases the regression model was:

$$y_{ij} = \mu + \alpha_i + \beta_1 x_{ij} + \beta_2 m_d + e_{ij}$$

where $\beta_1$is the regression coefficient for the direct additive effect of an animal i, x_{ij} is its EBV, β_2 is the regression coefficient for the maternal EBV and m_d is the EBV of the dam (d) of the selection candidate i. For model ME, the regression did not account for the permanent effect of the dam, due to the confusion it would cause between maternal EBVs and the dam effect.

The estimates of the regression coefficients were obtained from GenStat (Payne et al., 2009).

The accuracy of prediction was calculated as the correlation between the predicted EBV and actual phenotype. The correlation represents the accuracy of phenotype prediction (r_P). To obtain an accuracy of breeding value prediction (r_A), r_P was divided by the square root of the heritability of the trait. The estimate of h^2 was taken from the same analysis on basis of which the EBVs were calculated. The values used for the calculation of the correlation were obtained by correcting both the phenotype and the EBV for the fixed effects, i.e. the correlation was calculated between the residuals of these two variables. The accuracies were calculated for direct breeding value only, i.e. no accuracies of maternal breeding values were obtained.

4. RESULTS

4.1 PEDIGREE DEPTH AND NUMBER OF RECORDS

4.1.1 MODEL CHOICE ON FULL DATA

Table 3.4 presents the variance estimates and related parameters for BWT and HHP, calculated from the full dataset. Based on LRT, the best models for BWT and HHP were AcM and A respectively.

For BWT, inclusion of maternal effects, be it genetic or environmental, caused a significant reduction of the estimate of genetic variance, as compared to a simple animal model (A). The more effects were fitted in the model, the more of a reduction could be observed, i.e. inclusion of one maternal effect reduced the estimate of genetic variance by ~30% (for models ideM and M), while reduction in the AcM model was 39%. Although a proportion of the difference was distributed between the estimates of the maternal effects, models containing maternal effects had also increased error variance. The pattern of change was similar to that of the differences observed for σ_A^2, although the magnitude of the change was smaller, at the maximum increase of 22% for AcM. As a result, the estimates of the heritability decreased with increasing number of effects used. The estimates of heritabilities for the maternal effects (m^2 and *p.e.*) were of similar magnitude, at 0.02 (SE 0.00), with maternal genetic effects having marginally higher estimates than the permanent environment of the dam. Despite the small magnitude, the maternal effects were found to be significant, with standard error estimated very close to 0. The correlation between maternal genetic effect and direct genetic effects was found to be significant as well, at 0.11 (SE 0.04).

In contrast, for HHP the simple model was found to be best fitting the data. Introducing maternal effects to the models fitted to HHP resulted in re-shuffling of the variance between particular effects, with patterns similar to those observed in BWT, i.e. reduction of the additive genetic variance estimate to accommodate maternal effects, with concurrent increase in error variance, however, the differences were smaller (up to 7% decrease in the estimate of genetic variance) and not significant.

4.1.2 CHANGES OF VARIANCE OVER TIME

Table 3.5 shows the effect of the number of generations on estimation of variance components. For BWT, the genetic variance estimated using early records (P1) was significantly higher than the equivalent from the latter period (P2). The estimate of the

heritability for this period, despite being calculated with maternal effects included in the model, agrees with the estimate of the heritability obtained from simple animal model (A) run on the whole dataset. The reduction in the estimate of the additive variance between the periods was as high as 41%. Fitting the full pedigree to the phenotypic data of the second period (P2_ALL) resulted in a slight increase of genetic variance. The error variance was lowest in P1, highest in P2 and intermediate in P2_ALL. The changes in the estimates of maternal effects over periods were negligible.

For HHP, the analysis carried out on the early data showed the lowest estimate of genetic variance, while the highest estimate was obtained using P2_ALL scenario. Overall, the phenotypic variance for this trait differed between periods, with the more recent records covering a wider distribution, thus resulting in a larger estimate of the variance. The expansion of the variance in P2 was observed across variance components, i.e. both the additive and error variance were smaller in P1, with the effect particularly well defined for the latter. However, the differences were of much smaller magnitude than in BWT, with the maximum of 14% increase between P1 and P2. Due to large standard errors those differences between estimates for HHP were not significant.

The bivariate analyses performed on the data revealed a moderately strong negative correlation (r_A = -0.23, SE 0.04) between the genetic effects for the two traits, as shown in Table 3.6. The magnitude of the correlation was slightly higher for P1, at r = -0.25 (SE 0.03), although the difference was not significant. The patterns of change with varying data content followed the patterns found in univariate analyses. The estimates of variance components for BWT obtained from the bivariate analysis show very little departure from estimates obtained in univariate analyses of this trait. For HHP, fitting this trait together with BWT resulted in higher estimates of the genetic variance and marginally lowered error variance, however these differences were not statistically significant due to large standard errors. The bivariate analysis revealed also a significant negative residual correlation between traits.

Table 3.4 Estimates of variance components and corresponding heritabilities for BWT and HHP, based on univariate analysis of 1.3 million records. σ^2_A - variance of the direct genetic effect of an individual, σ^2_M - variance of the maternal genetic effect, $\sigma^2_{p.e.}$ - variance of the permanent environment effect of the dam, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2- direct heritability, m^2 - maternal heritability, p.e. - proportion of the total variance explained by permanent environment, r_{AM} - correlation between direct and maternal genetic effects. The best model identified through LRT is underlined.

Trait	Model	σ^2_A (SE)	σ^2_M (SE)	$\sigma^2_{p.e.}$ (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)	m^2 (SE)	*p.e.* (SE)	r_{AM} (SE)
BWT	A	148.19 (1.37)	-	-	116.93 (0.67)	265.12 (0.78)	0.56 (0.00)	-	-	-
	M	103.59 (1.67)	13.34 (0.39)	-	136.65 (0.81)	253.58 (0.82)	0.41 (0.01)	0.05 (0.00)	-	-
	ideM	101.09 (1.61)	-	7.66 (0.20)	137.80 (0.79)	246.55 (0.82)	0.41 (0.01)	-	0.03 (0.00)	-
	ME	94.39 (1.63)	5.86 (0.44)	4.76 (0.26)	140.90 (0.80)	245.91 (0.82)	0.38 (0.01)	0.02 (0.00)	0.02 (0.00)	-
	AcM	90.39 (2.00)	5.61 (0.44)	4.37 (0.27)	142.79 (0.97)	245.68 (0.82)	0.37 (0.01)	0.02 (0.00)	0.02 (0.00)	0.11 (0.04)
HHP	A	151.00 (9.13)	-	-	422.79 (6.56)	573.79 (6.00)	0.26 (0.01)	-	-	-
	M	144.56 (9.96)	3.13 (2.64)	-	425.05 (6.64)	572.74 (5.98)	0.25 (0.02)	0.01 (0.00)	-	-
	ideM	144.16 (9.91)	-	2.98 (2.54)	424.72 (6.55)	571.85 (6.02)	0.25 (0.02)	-	0.01 (0.00)	-
	ME	142.17 (10.16)	2.04 (2.90)	2.01 (2.91)	425.54 (6.62)	571.76 (6.01)	0.25 (0.02)	0.00 (0.01)	0.00 (0.01)	-
	AcM	139.61 (12.18)	1.64 (3.04)	1.83 (2.91)	426.77 (7.36)	571.92 (6.02)	0.24 (0.02)	0.00 (0.01)	0.00 (0.01)	0.14 (0.38)

Table 3.5 Estimates of variance components and corresponding heritabilities for BWT and HHP, based on univariate analysis of partial records. Set P1 included the pedigree and phenotypes of the first 12 generations, P2 contained the information for the second part. P2_ALL contained phenotypes for the latter generations, but utilized the full pedigree. σ^2_A - variance of the direct genetic effect of an individual, σ^2_M - variance of the maternal genetic effect, $\sigma^2_{p.e.}$ - variance of the permanent environment effect of the dam, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2- direct heritability, m^2 - maternal heritability, *p.e.* - proportion of the total variance explained by permanent environment.

Trait	Model	σ^2_A (SE)	σ^2_M (SE)	$\sigma^2_{p.e.}$ (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)	m^2 (SE)	*p.e.* (SE)
BWT	P1	130.55 (2.38)	4.35 (0.51)	3.17 (0.31)	112.89 (1.19)	250.96 (1.23)	0.52 (0.01)	0.02 (0.00)	0.01 (0.00)
	P2	76.14 (2.18)	5.69 (0.62)	6.11 (0.39)	156.90 (1.10)	244.84 (1.05)	0.31 (0.01)	0.02 (0.00)	0.02 (0.00)
	P2_ALL	83.30 (2.37)	6.48 (0.69)	5.83 (0.39)	155.61 (1.13)	251.22 (1.20)	0.33 (0.01)	0.03 (0.00)	0.02 (0.00)
HHP	P1	122.58 (12.27)	-	-	411.56 (9.39)	543.14 (8.01)	0.23 (0.02)	-	-
	P2	130.61 (14.02)	-	-	456.56 (10.88)	587.17 (9.42)	0.22 (0.02)	-	-
	P2_ALL	139.83 (15.03)	-	-	455.68 (10.93)	595.51 (10.04)	0.23 (0.02)	-	-

Table 3.6 Estimates of variance components and corresponding heritabilities for BWT and HHP, based on bivariate analysis of partial records. Set P1 included the pedigree and phenotypes of the first 12 generations, P2 contained the information for the second part. P2_ALL contained phenotypes for the latter generations, but utilized the full pedigree. σ^2_A - variance of the direct genetic effect of an individual, σ^2_M - variance of the maternal genetic effect, $\sigma^2_{p.e.}$ - variance of the permanent environment effect of the dam, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2- direct heritability, m^2 - maternal heritability, *p.e.* - proportion of the total variance explained by permanent environment, r_{AM} - correlation between direct and maternal genetic effects. r_A - correlation between direct genetic effects for the two traits, r_E - correlation between residuals.

Data set	Trait	σ^2_A (SE)	σ^2_M (SE)	$\sigma^2_{p.e.}$ (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)	m^2 (SE)	*p.e.* (SE)	*Correlations*
P1	BWT	130.75 (2.38)	4.28 (0.51)	3.18 (0.31)	112.79 (1.19)	251.00 (1.23)	0.52 (0.01)	0.02 (0.00)	0.01 (0.00)	r_A = -0.25 (0.03)
	HHP	135.99 (11.33)	-	-	408.73 (8.66)	544.72 (7.61)	0.25 (0.02)	-	-	r_E = -0.08 (0.01)
P2	BWT	76.19 (2.18)	5.61 (0.61)	6.14 (0.39)	156.88 (1.10)	244.81 (1.05)	0.31 (0.01)	0.02 (0.00)	0.03 (0.00)	r_A = -0.23 (0.03)
	HHP	137.90 (12.21)	-	-	460.42 (9.54)	598.32 (8.51)	0.23 (0.02)	-	-	r_E = -0.10 (0.01)
P2_ALL	BWT	83.41 (2.37)	6.38 (0.68)	5.86 (0.39)	155.56 (1.13)	251.21 (1.20)	0.33 (0.01)	0.03 (0.00)	0.02 (0.00)	r_A = -0.23 (0.03)
	HHP	147.88 (13.12)	-	-	459.37 (9.59)	607.24 (9.05)	0.24 (0.02)	-	-	r_E = -0.10 (0.01)

Table 3.7 Estimates of heritabilities for BWT obtained from fitting phenotypic records collected over 2-year periods and a full pedigree. σ^2_A - variance of the direct genetic effect of an individual, σ^2_M - variance of the maternal genetic effect, $\sigma^2_{p.e.}$ - variance of the permanent environment effect of the dam, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2- direct heritability, m^2 - maternal heritability, *p.e.* - proportion of the total variance explained by permanent environment, *N*- number of phenotypic records in the period, *F* - average inbreeding for the group.

Data set	σ^2_A (SE)	σ^2_M (SE)	$\sigma^2_{p.e.}$ (SE)	σ^2_e (SE)	h^2 (SE)	m^2 (SE)	*p.e.* (SE)	*N*	*F*
DAT_{Y1}	93.54 (5.65)	4.00 (1.50)	4.85 (1.17)	176.94 (2.98)	0.33 (0.02)	0.01 (0.01)	0.02 (0.00)	106,223	0.03%
DAT_{Y2}	95.63 (4.69)	3.54 (1.03)	5.27 (0.74)	141.67 (2.40)	0.39 (0.02)	0.01 (0.00)	0.02 (0.00)	173,040	1.31%
DAT_{Y3}	85.77 (4.06)	4.76 (1.02)	4.31 (2.04)	111.73 (2.04)	0.42 (0.02)	0.02 (0.00)	0.02 (0.00)	183,827	3.09%
DAT_{Y4}	78.56 (4.04)	4.05 (0.99)	5.49 (0.67)	124.01 (2.01)	0.37 (0.02)	0.02 (0.00)	0.03 (0.00)	178,491	4.62%
DAT_{Y5}	86.13 (4.27)	3.55 (0.96)	5.60 (0.66)	125.69 (2.08)	0.39 (0.02)	0.02 (0.00)	0.03 (0.00)	194,176	5.83%
DAT_{Y6}	71.79 (4.46)	6.40 (1.36)	5.34 (0.80)	173.10 (2.18)	0.28 (0.02)	0.02 (0.01)	0.02 (0.00)	175,610	8.47%
DAT_{Y7}	78.19 (4.99)	5.77 (1.43)	7.10 (0.88)	174.98 (2.35)	0.29 (0.02)	0.02 (0.01)	0.03 (0.00)	172,369	10.94%
DAT_{Y8}	66.91 (5.60)	11.96 (2.15)	6.62 (1.14)	166.01 (2.59)	0.27 (0.02)	0.02 (0.01)	0.03 (0.00)	119,771	12.84%

Table 3.8 Estimates of heritabilities for HHP obtained from fitting phenotypic records collected over 2-year periods and a full pedigree. . σ_A^2 - variance of the direct genetic effect of an individual, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance, h^2- direct heritability, *N*- number of phenotypic records in the period, *F* - average inbreeding for the group.

Data set	σ_A^2 (SE)	σ_e^2 (SE)	h^2 (SE)	*N*	*F*
DAT_{Y1}	88.07 (20.78)	516.16 (20.60)	0.15 (0.03)	2,991	0.03%
DAT_{Y2}	141.33 (21.93)	401.22 (17.02)	0.26 (0.04)	4,054	1.31%
DAT_{Y3}	117.65 (19.12)	351.05 (14.77)	0.25 (0.04)	3,918	3.09%
DAT_{Y4}	122.12 (20.28)	409.54 (16.24)	0.23 (0.03)	3,909	4.62%
DAT_{Y5}	144.98 (21.62)	369.18 (15.72)	0.28 (0.04)	4,100	5.83%
DAT_{Y6}	88.08 (19.71)	491.86 (17.11)	0.15 (0.03)	3,912	8.47%
DAT_{Y7}	228.15 (35.09)	520.241 (23.41)	0.30 (0.04)	3,918	10.94%
DAT_{Y8}	135.77 (30.34)	407.85 (22.59)	0.25 (0.05)	2,062	12.84%

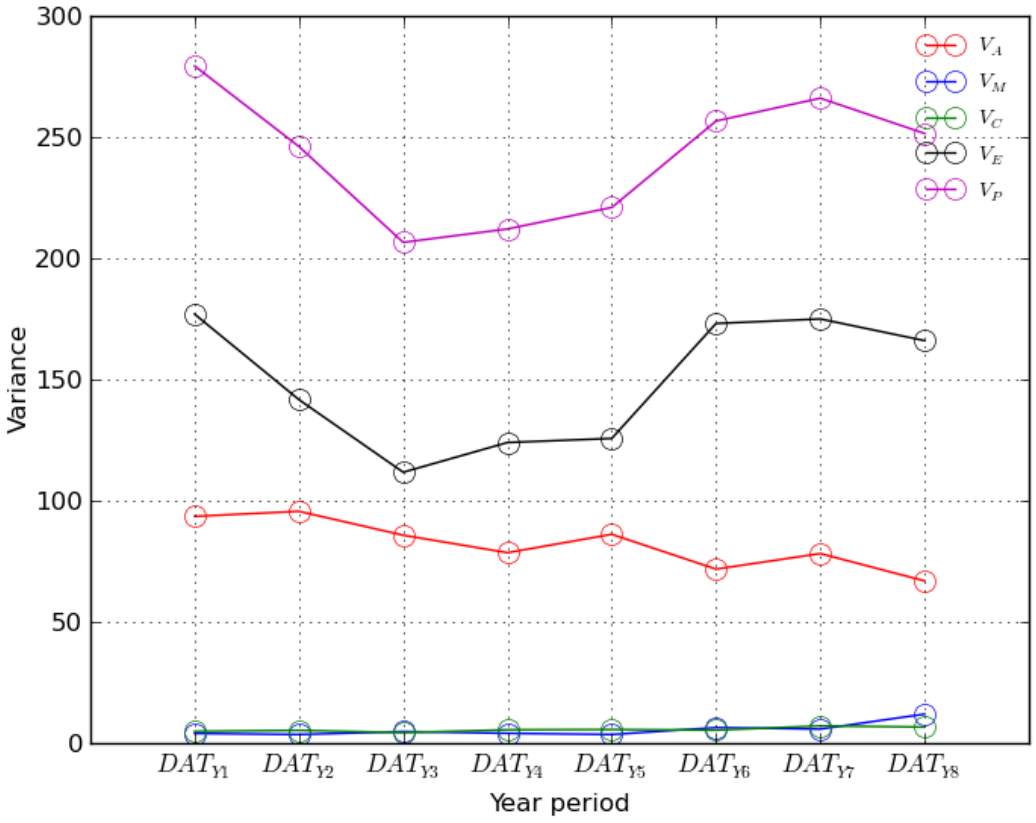

Figure 3.1 Changes of BWT variance component estimates in consecutive periods of phenotype recording

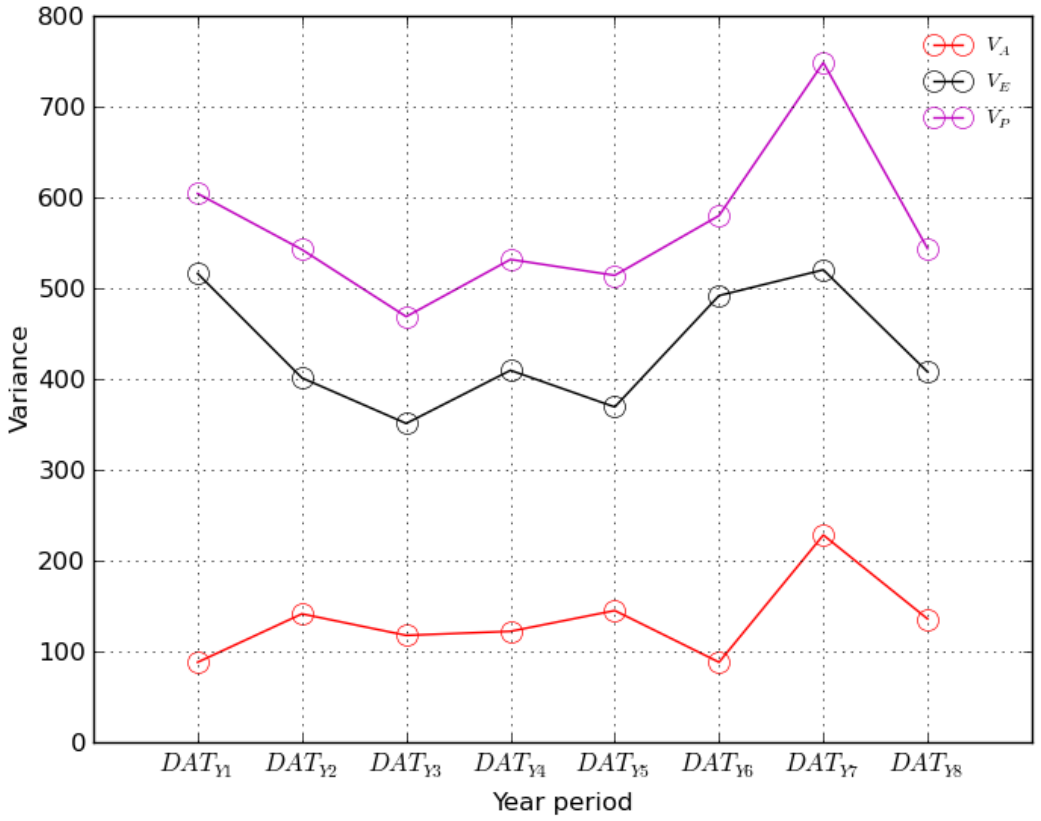

Figure 3.2 Changes of HHP variance component estimates in consecutive periods of phenotype recording

Table 3.7 provides further information on the effect of splitting the phenotypic records into smaller periods. The estimates of genetic variance were found to decrease gradually over the two year periods. The differences between neighboring periods were not significantly different, however, comparing the earliest estimates with the latest yielded a significant reduction (28%) in the estimate of the variance. The estimates of the maternal variance did not show a clear pattern of change for the initial recording periods, with a sudden increase in the most recent records. This increase was significant for maternal genetic variance, but not for the permanent environment of the dam. The pattern of change in the estimates of error variance shows three distinct parts. Error variance estimates showed initial decrease, followed by stabilization and finally an increase back to values approximating the initial estimates. The lowest estimate of the error variance showed a 37% reduction as compared to the highest value estimated from the first period of observations. The patterns of change observed in the estimates of variance are presented graphically in Figure 3.1.

Table 3.8 and Figure 3.2 present the equivalent changes of variance for HHP. Due to much lower numbers of records available, the majority of the differences in genetic variance observed between periods are not statistically significant, with the exception of the most extreme values found for DAT_{Y1}, DAT_{Y6} and DAT_{Y7}. There is no clear pattern of change both for the genetic and environmental variance for this trait.

4.2 Bias and Accuracy of Predictions

4.2.1 Univariate Analysis

Removing the phenotypic records of the last two generations resulted in slightly increased estimates of heritability, both for BWT and HHP, as shown in Table 3.9. The EBVs predicted for the selection candidates showed an overestimation of the range of the true BVs (TBVs) for BWT, and underestimation for HHP EBVs. For BWT, the simple model produced estimates with largest bias. Adding either environmental or maternal genetic effect decreased the bias significantly, with the reduction larger in ideM than in the M model. Incorporating both types of maternal effects in the ME model resulted in the same optimum bias of the direct EBV as in ideM model, but also presented the improved estimation of the maternal EBV. The accuracy of BWT phenotype predictions did not differ between models and was estimated at 0.16 (SE 0.004). After scaling it by the $\sqrt{h^2}$, with the estimate obtained from the appropriate model, the accuracy of BWT EBVs predicted from the simple animal model was significantly worse than models including maternal terms, but the differences between maternal models were negligible. The best accuracy of BWT EBVs was thus obtained from models ideM and ME, at 0.25 (SE 0.006).

The bias of HHP predictions from tested models showed an opposite effect to BWT, with regression coefficient larger than 1 suggesting underestimation of the range of TBVs. However, due to large standard errors the differences between the bias estimates for the models were not statistically different. The accuracy of HHP prediction was constant between models, and higher than accuracy of BWT predictions. Accuracy of HHP phenotype prediction was estimated at 0.21 (SE 0.02) and accuracy of breeding value prediction was estimated at 0.42 (SE 0.05).

4.2.2 Bivariate Analysis

Table 3.10 shows the bias and accuracy of BWT and HHP predictions, based on bivariate analysis. The results showed little improvement over the univariate analyses, with the exception of the prediction of the direct EBV for BWT, which was less biased in bivariate,

than in univariate analyses. Keeping the BWT records for selection candidates (S2) improved the accuracy of HHP EBV predictions, without changing the bias.

Table 3.9 The estimates of bias in univariate analyses, expressed as regression coefficients of recorded phenotypes on predicted BVs using different models and of accuracy, calculated as a correlation between residuals left after correcting phenotypes and EBVs for fixed effect levels. β_A - bias of direct EBVs, β_M - bias of maternal EBVs, r_P - accuracy of phenotype prediction, r_{EBV} - accuracy of breeding value prediction.

Trait	Model	h^2 (SE)	m^2 (SE)	β_A (SE)	β_M (SE)	r_P (SE)	r_{EBV} (SE)
BWT	A	0.57 (0.00)	-	0.66 (0.02)	-	0.16 (0.004)	0.21 (0.006)
	ideM	0.43 (0.01)	-	0.85 (0.02)	-	0.16 (0.004)	0.25 (0.006)
	M	0.43 (0.01)	0.05 (0.00)	0.79 (0.02)	0.66 (0.04)	0.16 (0.004)	0.24 (0.006)
	ME	0.41 (0.01)	0.02 (0.00)	0.85 (0.02)	0.87 (0.07)	0.16 (0.004)	0.25 (0.006)
HHP	A	0.26 (0.01)	-	1.17 (0.13)	-	0.21 (0.02)	0.42 (0.05)
	ideM	0.25 (0.02)	-	1.22 (0.14)	-	0.21 (0.02)	0.42 (0.05)
	M	0.25 (0.02)	0.01 (0.00)	1.31 (0.15)	-3.43 (2.49)	0.21 (0.02)	0.41 (0.05)
	ME	0.25 (0.02)	0.00 (0.01)	1.32 (0.15)	-5.41 (3.87)	0.21 (0.02)	0.41 (0.05)

Table 3.10 The estimates of bias in bivariate analyses, expressed as regression coefficients of recorded phenotypes on predicted BVs using different models and of accuracy, calculated as a correlation between residuals left after correcting phenotypes and EBVs for fixed effect levels. S1 - prediction of BWT and HHP EBVs with records of both traits masked in the selection candidates, S2 - prediction of HHP EBVs with BWT records of selection candidates kept in the data; β_A - bias of direct EBVs, β_M - bias of maternal EBVs, r_P - accuracy of phenotype prediction, r_{EBV} - accuracy of breeding value prediction.

Analysis	Trait	h^2 (SE)	m^2 (SE)	β_A (SE)	β_M (SE)	r_P (SE)	r_{EBV} (SE)
S1	BWT	0.41 (0.01)	0.02 (0.00)	0.86 (0.02)	0.85 (0.07)	0.17 (0.00)	0.26 (0.01)
	HHP	0.25 (0.01)	-	1.17 (0.14)	-	0.20 (0.03)	0.41 (0.06)
S2	HHP	0.25 (0.01)	-	1.18 (0.13)	-	0.21 (0.03)	0.43 (0.06)

5. DISCUSSION

The analyses presented in this Chapter were based on a vast dataset, far exceeding the largest studies previously published, at 1.3M broiler birds versus 570K used in the previous largest study (Koerhuis and McKay, 1996). The wealth of the data used in the presented study is not only expressed in the numbers of the phenotypic records available, but also in the number of generations used, reaching up to 24. This depth of the pedigree allowed a thorough examination of the changes of variance in BWT and HHP over the last 15 years. Considering the spectacular progress achieved in broiler breeding programs, such evaluation of changes in variance over time provides an invaluable insight into the amount of genetic variability and the ability of BLUP predictions to account for the effects of selection.

5.1 MODEL CHOICE

The presented analysis of the large number of BWT records provides solid evidence for the presence of the maternal genetic effects acting on this trait. Abundance of the data allowed a more precise examination of the correlation between direct and maternal genetic effects than could be achieved before, with the estimate of r_{AM}=0.11 (SE 0.04) standing in contrast with previous findings of Chapter 2 and other, published studies (Koerhuis et al., 1997, Koerhuis and Thompson, 1997). It is difficult to explain the discrepancy between the results presented in this analysis and the results of the study by Koerhuis and Thompson (1997), however, the positive sign of this correlation is supported by the strong and positive correlation between the juvenile body weight and egg weight, with the latter believed to represent the maternal effect on the weight of the chick (Koerhuis and Thompson, 1997, Pakdel et al., 2002).

One of the major limiting factors in the routine use of the comprehensive models including multiple random terms, together with their covariances, is the large computational time and power required for their convergence. For example, the univariate analysis of the 1.3M records for BWT using the AcM model took 33.5h to converge. As the commercial evaluations are based on multivariate analyses, routine use of such extended models may not prove feasible. However, the apparent inconvenience of such approach needs to be weighed carefully against the benefits obtained from inclusion of all significant sources of variation. In the presented results, the estimates of the genetic variance obtained from the simple model (A) seemed to be unrealistically high, with the estimate of the genetic variance 64% larger than the estimate obtained from the most comprehensive model. Based on simulation study, omission of significant maternal effects in the models has been shown to inflate the estimates of the direct genetic variance, with the maternal influence mistakenly treated as the additive

effect that the dam passes to her offspring (Clément et al., 2001). This inflated estimate of genetic variance leads to an overestimation of the range of the TBVs and thus introduces bias into EBV predictions (Clément et al., 2001).

5.2 BIAS OF PREDICTIONS

5.2.1 BWT PREDICTIONS

The result of this study prove the expectations of bias derived from simulation studies in the real data, with the EBVs predicted from models accounting for maternal effects suffering less bias than the simplistic model A. However, this reduction in bias can also have other explanations. For example, maternal effects could be explaining part of the selection bias not accounted for by BLUP corrections. Although the BWT records are used as a core trait for the first stage of the selection, there is a strong possibility that extremely poor birds are excluded from the growing stock before their BWT records are collected. As maternal effects are expected to exert the largest influence in very young animals, it is possible that including them in the model used for prediction of BWT EBVs accounts for at least a part of this "pre-selection" bias. This would of course mean that there is indirect selection acting on the maternal effects, thus inducing maternal selection bias. This would agree with the results found here, with the regression coefficient for maternal EBVs being less than 1. However, the bias of the maternal EBVs could be also explained by the positive correlation between the direct and maternal effects observed in the AcM model, which was not accounted for in the model used for predictions (ME).

While the maternal effects coped with some of the bias affecting the predictions for BWT, none of the models reached the unbiasedness expected for BLUP, i.e. $\beta=1$. The estimates of bias presented in this study are, to my knowledge, the only estimates available for BLUP prediction bias in broilers. However, results from layers (Wolc et al., 2011) show that the assumption of unbiased EBV prediction via BLUP rarely holds in real layer populations. In practice, there are many factors influencing the changes in genetic variance in a population, not many of which can be formally accounted for. In the studied example of BWT, the regression coefficient consistently lower than 1 indicates that the estimate of the genetic variance used for predictions is inflated, as compared to the variance present in the trait. Considering the heritability estimate used in the predictions has been calculated on the basis of 24 generations, this overestimation is hardly surprising, however, it highlights the importance of the use of appropriate estimates of genetic variance used for predictions. This effect is particularly pertinent to models which do not account for maternal effects, as the

inflation of the estimates of the genetic variance in these is magnified through incorrect assignment of the observed variance.

Inclusion of maternal effects in the models used for predictions creates an opportunity to select for the maternal EBVs. As can be seen from the regression coefficients, these maternal EBVs are estimated with similar precision as the direct EBVs, thus it is not unreasonable to expect that progress achieved by selecting for direct EBVs could be replicated for the maternal traits. Although the magnitude of the maternal effects for BWT is consistently small between studies (e.g. Koerhuis and Thompson, 1997, Pakdel et al., 2002), the maternal influences are considered as associative effects, as discussed in Chapter 1, thus their impact will be much larger than would be expected for individual-oriented terms of the same magnitude.

5.2.2 HHP PREDICTIONS

In contrast to BWT, the maternal effects seem to have no effect on the HHP. This stands in agreement with the widespread belief that the influence of the dam is expressed mostly in the early life of the individual, with the traits exerted in adults determined mostly by the additive effect of the individual itself (as reviewed by Lande and Kirkpatrick, 1990). The interpretation of the re-shuffling of the variance with the different models fitted to HHP was hindered by the large standard errors, however the findings provide some support to the simulation results obtained by Clément et al. (2001). These showed that the estimation of the genetic variance is not hindered by inclusion of maternal effects in a model, even when the trait analysed is not affected by them.

The lack of significant differences between models and lack of significant maternal effects for HHP in the presented study were also extended to the estimates of the bias for this trait. The regression coefficients for the direct effect of HHP were consistently indicating an underestimation of the range of true TBVs for this trait. The selection based on BWT could explain the bias of univariate HHP predictions, however, the lack of marked improvement in the bivariate analysis of the two traits indicates other sources of the reduction in HHP variance, possibly caused by selection on other traits. As the HHP records are collected on adult birds, the effect of natural selection is also likely to affect the variance of this trait. Although mortality records have been introduced into routine evaluations of breeding stock, they are still only a poor approximation of fitness of the breeding birds, where the same score is given to birds which died as a result of an accident, i.e. broken leg, and birds that died for unexplained reasons. In practice, culling based on fitness traits is largely

unaccounted for, thus introducing bias into predictions across production traits (Quinton, 2003).

Interestingly, the accuracy of predictions for HHP exceeded the accuracy of BWT prediction, with the accuracy of breeding value prediction reaching values of 0.4 and 0.25 for HHP and BWT respectively. The accuracy of prediction is expected to be larger for traits with higher heritability. Considering the direction of the bias for BWT it appears that using inflated estimate of the genetic variance for this trait reduces the accuracy of predictions. This reduction is not observed in HHP, as the estimate of variance for this trait has not changed over time.

5.3 Changes of Variance over Time

Considering the high accuracy of the HHP predictions, it would be expected that selection pressure applied to this trait would reduce the genetic variance over time. However, the estimates of the σ_A^2 obtained from the early and late periods (P1 and P2) show very limited changes for this component, with the estimates obtained from the more recent period apparently exceeding the estimates from early generations. As shown in the examination of the two year periods, the increased estimate of the variance for the latter period is likely due to a single high estimate of DAT_{Y7}. Due to large standard errors of the estimates it is difficult to interpret the results, however, it is likely that due to truncation of the available phenotypes as a result of the juvenile selection based on growth and feed efficiency, the true variance is higher than would be expected from routine evaluations. The selection is assumed to reduce the genetic variance, thus BLUP treats the variance of the phenotypic records of P2 as a reduced proportion of the full variance present in the base. Therefore, fitting the same phenotypic records, with a larger pedigree in P2_ALL resulted in a higher estimate of the variance.

The maintained genetic variance of HHP most likely indicates that the selection pressure applied to this trait is not strong, despite the high accuracy with which the EBVs are predicted. This result is not surprising considering the negative correlation of this trait with BWT (r=-0.25, SE 0.03) and other growth and feed efficiency traits (as reviewed by Chambers, 1990). While the adult stage of the selection favours birds with high egg production, the juvenile selection based on growth traits is likely to select birds excelling in those traits that would also rank the birds lower for HHP. Thus, the selection applied to HHP can be seen as a stabilizing selection, explaining the maintained variance.

As outlined in the introduction, the estimates of the variance used for EBV prediction in BLUP accommodate for the effects of selection exerted through the Bulmer effect and inbreeding. The levels of inbreeding observed in this population are increasing steadily, below the maximum of 1% per generation threshold outlined by FAO (2010). The BLUP correction for the inbreeding effects can be observed in the results of the P2_ALL analyses, however, its magnitude does not explain the full extent of the reduction observed between the periods. The estimate of the genetic variance obtained in the P2_ALL analysis is only 9% higher than the estimate of P2 and is 36% lower than the P1 estimate, which in this case is treated as the reference. This shows that the BLUP correction for the effects of inbreeding is not adequate to the magnitude of the changes observed.

It could be argued that splitting the phenotypic data into 2-year periods makes the correction for the Bulmer effect impossible, as it requires the full information for the population under selection (Bulmer, 1971). Thus, it could be speculated that the reduction of variance that was not accounted for by the correction due to inbreeding could be explained by the Bulmer effect. As such, routine evaluations which, in general, use the continuous data could be more accurate as they may account for the Bulmer effect. However, the Bulmer effect is largest in the first generation of selection (Bulmer, 1971). The pedigrees used in the commercial evaluations are inevitably breaking the assumption of the unselected base population, therefore the actual extent of this effect in livestock populations is questionable.

Considering the number of generations used, it is likely that a proportion of the changes in the variance observed is caused by the directional effect of selection on allele frequencies, which is not accounted for by the infinitesimal model. While these effects are assumed to be negligible over short periods (Bulmer, 1971), their cumulative effect after 20 generations could be non-trivial.

5.4 Environmental Changes

The analysis of the change in variance components over time provides also insight into the environmental influences on the phenotypes. For both BWT and HHP, the pattern of change of the residual variance reflects the management changes that affected the broiler production, which could be explained by two major changes in the environment of the poultry stock. Around the time of the most drastic change in the estimates of the residual variance, i.e. DAT_{Y6} and DAT_{Y7}, the diet of Aviagen broilers changed dramatically, with the fish meal removed from the standard diet composition (A. Kranis, personal communication). In addition, this period coincided with the introduction of the ban on use of antibiotics in feed

as growth promoters within the EU (EC, 2005). Such significant changes in the diet are likely to be reflected in the noise of the estimates over a learning period during which the management adapts to the new conditions.

5.5 Bivariate Predictions

Due to the experimental nature and extensive computational requirements, the analyses presented in this Chapter were limited to two traits only. However, it is expected that multivariate analyses would increase the precision of both variance component estimation and accuracy of predictions, by providing more information on the sources of selection bias (Mrode, 2005). The bivariate model using BWT and HHP records presented in this Chapter provides some preliminary evidence for this, with the estimation of the genetic variance for HHP showing more precision when analysed together with BWT than in univariate analyses. As was expected, the benefits of fitting the traits together were more visible for HHP than BWT, as the HHP is a trait of lower heritability and is recorded after the juvenile selection. Interestingly, while the benefits of this approach were obvious in the estimation of variance, they were not clear in the precision of the EBV predictions, with no significant differences between bias and accuracy of predictions based on uni- and bi-variate models.

6. Conclusions

The reduction of the genetic variance in BWT as a result of selection acting over the last 15 years is considerable and cannot be explained by inbreeding rates alone. The inflation of the genetic variance estimate can be also caused by using simplistic models which do not account for significant maternal effects. Using inflated estimates of this variance in EBV predictions leads to reduced accuracy and increased bias. The effect of selection seems to over-ride the dependence of accuracy on heritability, with HHP, a trait with lower heritability than BWT, achieving higher accuracies of predictions. The maintained genetic variance of HHP indicates that this trait is either subject to very weak selection pressure, or that the divergent goals of selection stages have a stabilizing effect on this trait. The results of this study highlight the need for a balance between the benefits of deep pedigrees used for estimation of the significant effects for a trait, and the risks of using estimates which are not pertinent to the current populations.

7. REFERENCES

ARTHUR, J. A. & ALBERS, G. A. 2003. Industrial Perspective on Problems and Issues Associated with Poultry Breeding. *In:* MUIR, W. M. & AGGREY, S. E. (eds.) *Poultry Genetics, Breeding and Biotechnology.* CABI Publishing.

BULMER, M. G. 1971. The Effect of Selection on Genetic Variability. *The American Naturalist,* 105**,** 201-211.

BULMER, M. G. 1980. *The mathematical theory of quantitative genetics*, Clarendon Press.

CHAMBERS, J. 1990. Genetics of Growth and Meat Production in Chickens. *In:* CRAWFORD, R. (ed.) *Poultry Breeding and Genetics.* Elsevier.

CLÉMENT, V., BIBÉ, B., VERRIER, É., ELSEN, J.-M., MANFREDI, E., BOUIX, J. & HANOCQ, É. 2001. Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects. *Genetics Selection Evolution,* 33**,** 369-396.

CUTTER, A. D. & PAYSEUR, B. A. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet,* 14**,** 262-274.

DARWIN, C. 1859. *On the Origin of Species by means of Natural Selection, or the preservation of favoured races in the struggle for life,* London, John Murray.

FALCONER, D. S. & MACKAY, T. F. C. 1996. *Introduction to quantitative genetics / D.S. Falconer and Trudy F.C. Mackay*, Harlow : Longman, 1996.4th ed.

FAO, F. A. A. O. O. T. U. N. 2010. BREEDING STRATEGIES FOR SUSTAINABLE MANAGEMENT OF ANIMAL GENETIC RESOURCES.

GILMOUR, A. R., GOGEL, B. J., CULLIS, B. R., WELHAM, S. J. & THOMPSON, R. 2006. *{ASREML} User Guide Release 2.0*.

HENDERSON, C. R. 1950 Estimation of genetic parameters. Biometrics, 6, 186-187.

HENDERSON, C. R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,* 31**,** 423-447.

HENDERSON, C. R. 1985. Best Linear Unbiased Prediction Using Relationship Matrices Derived from Selected Base Populations. *Journal of Dairy Science,* 68**,** 443-448.

HENDERSON, C. R., KEMPTHORNE, O., SEARLE, S. R. & VON KROSIGK, C. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics,* 15**,** 192-218.

HOFER, A. 1998. Variance component estimation in animal breeding: a review. *Journal of Animal Breeding and Genetics,* 115**,** 247-265.

KOERHUIS, A. & THOMPSON, R. 1997. Models to estimate maternal effects for juvenile body weight in broiler chickens. *Genetics Selection Evolution,* 29**,** 225 - 249.

KOERHUIS, A. N. M. & MCKAY, J. C. 1996. Restricted maximum likelihood estimation of genetic parameters for egg production traits in relation to juvenile body weight in broiler chickens. *Livestock Production Science,* 46**,** 117-127.

KOERHUIS, A. N. M., MCKAY, J. C., HILL, W. G. & THOMPSON, R. 1997. A genetic analysis of egg quality traits and their maternal influence on offspring-parental regressions of juvenile body weight performance in broiler chickens. *Livestock Production Science,* 49**,** 203-215.

LANDE, R. & KIRKPATRICK, M. 1990. Selection Response in Traits with Maternal Inheritance. *Genetical Research,* 55**,** 189-197.

MRODE, R. A. 2005. *Linear models for the prediction of animal breeding values*, Wallingford: CABI.

PAKDEL, A., VAN ARENDONK, J., VEREIJKEN, A. & BOVENHUIS, H. 2002. Direct and maternal genetic effects for ascites-related traits in broilers. *Poultry Science,* 81**,** 1273-1279.

PAYNE, R. W., MURRAY, D. A., HARDING, S. A., BAIRD, D. B. & SOUTAR, D. M. 2009. GenStat for Windows (12th Edition) Introduction.

http://www.vsni.co.uk/downloads/genstat/release12/doc/IntroGuide.pdf: VSN International.
QUINTON, M. 2003. *Use of mixed model methodology in poultry breeding: assumptions, limitations anc concerns of BLUP-based selection programs. In:* MUIR, W. M. & AGGREY, S. E. (eds.) *Poultry Genetics, Breeding and Biotechnology.* Wallingford: CABI .
VAN DER WERF, J. H. & DE BOER, I. J. 1990. Estimation of additive genetic variance when base populations are selected. *Journal of Animal Science,* 68**,** 3124-32.
WOLC, A., STRICKER, C., ARANGO, J., SETTAR, P., FULTON, J., O'SULLIVAN, N., PREISINGER, R., HABIER, D., FERNANDO, R., GARRICK, D., LAMONT, S. & DEKKERS, J. 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution,* 43**,** 5.
WOOLLIAMS, J. A., WRAY, N. R. & THOMPSON, R. 1993. Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genetics Research,* 62**,** 231-242.

CHAPTER 4

ACCURACY AND BIAS IN GENOMIC AND PEDIGREE BASED BREEDING VALUE PREDICTIONS

1. INTRODUCTION

Genomic Selection (GS) is a method of artificial selection used by livestock breeders, in which selection candidates are evaluated with the use of genome-wide estimated breeding values (GEBVs), predicted with the use of marker genotypes (Meuwissen et al., 2001). The calculation of GEBVs has been termed genomic prediction (GP), with numerous methods of GP developed over the last decade (e.g. Meuwissen and Goddard, 2001, Calus et al., 2008, de Los Campos et al., 2013). One of the most frequently used methods is based on the traditionally used Best Linear Unbiased Prediction (BLUP), in which the numerator relationship matrix **A** is replaced with a realized relationship matrix **G** calculated from the genotypes. This approach is termed GBLUP, and it has proved to be an effective method for evaluation of traits which lack predominant QTLs of large effect (Daetwyler et al., 2010b).

Irrespective of the method used, the benefits of using GP over predictions based on pedigree and phenotype alone include an improved accuracy of predictions, the shortening of generation interval and a possibility for reduced costs of phenotype recording. These benefits are expected to be largest for traits with low heritability, such as mortality, traits which are difficult to measure on selection candidates, e.g. carcass traits or sex-limited traits, or traits recorded late in life, such as fertility (Meuwissen, 2003). For highly heritable production traits which can be estimated with high accuracy, i.e. milk production in cattle evaluated through Progeny Testing schemes, the benefits of GS revolve around shortened generation interval. By using marker information, highly accurate GEBVs can be obtained as early as at birth of the selection candidate, thus reducing the generation interval and minimizing costs incurred by rearing the candidate and possible offspring. As a result, GS is being routinely used in species like cattle (Pryce and Daetwyler, 2012), sheep (Daetwyler et al., 2010a) or pigs (Cleveland et al., 2012).

In contrast, the application of GS in poultry breeding is only in its infancy. Despite the fact that chicken is one of the first species for which a complete genome sequence was published (Wong et al., 2004) and that there is over 78 million known polymorphisms segregating in chicken genome (Kranis et al., 2013), experiments where molecular markers are used for

calculation of GEBVs in this species have been limited. In contrast to species like cattle, shortening of the generation interval obtained through use of GS has a limited impact on broilers, where the generation interval is already short, much earlier than the time needed for the birds to reach sexual maturity. Further, the costs of genotyping far outweigh the monetary value of the individual selection candidate, thus creating an economical challenge for the routine use of GS (Preisinger, 2012).

Most studies published on GS in broiler chickens (e.g. Abdollahi-Arpanahi et al., 2014, Chen et al., 2011), concentrated on typical production traits, which have moderately high heritability and are recorded early in life, such as body weight or breast muscle. Although those studies reported some benefits of GP over traditional predictions, these traits have been already successfully selected upon by traditional, pedigree based selection. However, it would be far more interesting to assess the efficacy of GP in reproductive traits. As these traits are usually sex-limited and show low heritability, an accurate prediction of EBVs using genotypes, i.e. available for animals of both sexes, could potentially revolutionise the progress achieved in selection for broiler reproduction.

Broiler breeder traits are usually characterised by low heritabilities and have a complex nature, with reproductive rates affected by genetic and environmental effects of both sire and dam (Wolc et al., 2009). They are also considered as challenging for phenotype recording, as measurements can only be collected after the selection candidates have reached sexual maturity and usually are recorded on one sex only. For example, fertility in poultry is defined as a percentage of eggs laid that were fertile. It has been traditionally considered as a trait of a male, partly due to the fact that one male is mated to several females, partly due to the fertilization depending on several characteristics of sperm, such as mobility and volume (Foote, 2003). However, recent studies revealed a significant effect of the female on the fertilization success, albeit explaining less variance than that of males, with corresponding heritabilities of weekly fertility rates estimated at 7-11% for males and 7-9% for females (Wolc et al., 2009). Similarly, hatchability of broiler chickens has been found to be a complex trait, influenced by the effects of both dam and sire, although in contrast to fertility, the effect of the dam is larger than that of the sire and is primarily expressed through the dam's egg quality traits (Wolc and Olori, 2009).

Another example of a trait which is expected to benefit most from use of GS is mortality, a typical fitness related trait. Although birds with high survival rates are necessary for continuity of improvement programmes and for high levels of production, this trait, or rather a combination of traits, is notoriously difficult to improve, due to low levels of genetic

variance identified by traditional methods, with heritability depending on the age of recording and specificity of measurement. For example, heritability of general mortality recorded between 14 and 42 days of age has been estimated at 0.02 (SE 0.01) (Gonzalez-Recio et al., 2008) while heritability of mortality due to ascites-related traits was estimated at 0.22 (no SE given), with the component traits having lower heritabilities of 0.06 (mortality due to ascites), 0.10 (mortality due to heart failure), and 0.15 (mortality due to heart-lung failure) (de Greef et al., 2001). In poultry, mortality encompasses several different causes of death, e.g. due to ascites, heart failure (de Greef et al., 2001) or bacterial infection (Fossum et al., 2009). As such, the genetic determination of survival is likely to be very complex and difficult to estimate.

Thus, although broiler production has achieved spectacular success in improvement of meat production and feed efficiency, response to selection in reproductive and fitness traits using pedigree based predictions has been limited. Sourcing information directly from the genotypes of selection candidates could potentially overcome the limitations met by traditional methods of EBV prediction. The aim of this experiment was to assess the benefits of using marker information in prediction of EBVs for a range of difficult-to-measure traits in broiler chickens. Observed benefits were quantified by comparing accuracy and bias of genomic predictions at various marker densities to the reference of traditional predictions based on pedigree only.

2. MATERIALS

2.1 GENOTYPES

The dataset used in this analysis consisted of 5,416 birds genotyped using the full range of markers found on the 600k Affymetrix Axiom panel, spread over 28 chromosomes and sex chromosome Z. The dataset consisted of 1,446 birds genotyped using this high density panel, while the rest was genotyped using Illumina GoldenGate 3k SNP chip (Muir et al., 2008) and imputed to the full range of the Axiom panel. The imputation was performed by A. Kranis (Aviagen Ltd.) using AlphaImpute (Hickey et al., 2011). The genotypes were imputed for offspring of HD-genotyped sires and grandsires, which resulted in high accuracy (>0.97) between true and imputed genotypes in a validation subset.

2.2 PEDIGREE

Once initiated, the genotyping of individuals in this line was extensive. Thus, in the construction of the pedigree used in following analyses, the ungenotyped ancestors were considered unknown. The earliest genotyped individuals were therefore considered as the base population. The base population consisted of 206 genotyped individuals with both parents ungenotyped and assumed unknown, and 82 with either sire or dam unknown (8 with unknown sire when dam was known and 74 with unknown dam, when sire was known). Although the majority of the base population consisted of the chickens that hatched earliest, there were also a few genotyped individuals with unknown parents added at later stages. The genotyped individuals were spread over 7 generations. There were 320 sires which were offspring to 124 paternal grandsires and 165 paternal granddams and 1132 dams which were offspring to 179 maternal grandsires and 385 maternal granddams,. The average numbers of offspring were 16 (range 1 - 99) per sire and 5 (range 1 - 37) per dam.

2.3 PHENOTYPES:

Phenotypic measurements on 8 traits were analysed. Table 4.1 presents the trait description and number of records for each trait in training and test populations. Juvenile body weight was recorded on all individuals. Mortality was recorded on individuals of adult birds of both sexes, throughout the breeding period. Other traits were recorded for females only and are measures of the reproductive abilities of the hens. As such, they are recorded late in life, after the birds have passed all stages of selection.

To assess the predictive abilities of models used, the population was split into training (TRN) and test (TST) populations with approximately 60% individuals placed in TRN and remaining individuals treated as selection candidates. TRN contained individuals with both genotypes and phenotypes, while TST consisted of individuals treated as young selection candidates, i.e. with known genotypes but without phenotypic records. Records available for these individuals were masked during the breeding value prediction, and later used to compare predicted BVs with the observed phenotype. The choice of TST individuals was based on the hatch time, so that TST individuals were either offspring and/or siblings of TRN animals. None of the TST individuals had offspring in the TRN.

Table 4.1 Trait description, fixed effects used for given trait and numbers of records in training (TRN) and test populations (TST).

Trait	Description	Fixed Effects	TRN	TST	Total
BWT	Juvenile body weight at 35 days	hatch week + sex	3,162	2,254	5,416
HHP	Hen housed production	hatch week	2,469	533	3,002
EFERT	Early fertility - % of fertile eggs from start to 40 weeks	mate	2,490	445	2,935
LFERT	Late fertility – from 41 weeks on	late mate	2,420	314	2,734
EHOF	Early hatchability - % of fertile eggs that hatched up to 40 week	mate	2,479	443	2,922
LHOF	Late hatchability – from 41 weeks on	late mate	2,289	286	2,575
MORT	Mortality for the duration of laying period	hatch week	3,150	785	3,935

2.4 CHIP DENSITY

To assess the effect of marker density on accuracy of breeding value prediction, five chips were used in the analyses:

- 2k chip - containing markers present on proprietary 3k Illumina GoldenGate array (Muir et al., 2008), which passed quality control (QC) in this data set
- 42k chip - containing markers present on proprietary 60k Illumina BeadChip array (Groenen et al., 2011), which passed QC in this data
- 74k chip - an in-silico chip created by adding a random selection of markers to the 42k chip
- 175k chip - as above, an in-silico chip created by adding a random selection of markers to the 74k chip
- 431k chip - which consisted of markers from the complete set of the 600k Affymetrix Axiom chip that remained after QC in the presented data

Quality control was carried out after the markers were selected from the full marker set. Table 4.2 presents the number of markers available through the initial choice and proportion of SNPs that failed quality control performed in Plink (Purcell et al., 2007). The following markers were removed: markers not mapped to a known position, markers which failed Hardy Weinberg Equilibrium test with p<=0.001, markers with more than 5% genotypes

missing and markers with minor allele frequency MAF<0.01. While markers on Z chromosome were not removed in the quality control step they were not used for the calculation of genomic relationship matrices.

Table 4.2 Number of markers on each chip and proportion of markers that failed quality control due to defined criteria.

Chip name	Mapped markers	Markers failed due to:			Remaining
		HWE p<=0.001	Missing > 0.05	MAF<0.01	
2k	3,072	7%	2%	11%	2,483
42k	52,618	7%	3%	11%	41,757
74k	99,542	5%	3%	19%	73,620
175k	248,317	4%	4%	23%	174,718
431k	625,995	4%	4%	25%	431,249

3. METHODS

3.1 BREEDING VALUE PREDICTION

Breeding values of TST individuals were predicted based on their genotypes, using the phenotypes and genotypes of their relatives in TRN population and the relationships they shared. This was achieved by fitting mixed linear models in ACTA software package (Gray et al. 2012) for genomic predictions (GBLUP) and in ASReml for pedigree based predictions (Gilmour et al., 2006). Using two separate software packages was necessary due to the computational limitation of both ACTA and ASReml. At the time of the study, ACTA software allowed fitting pedigree based numerator relationship in the mixed linear models, however, in such cases it did not predict breeding values of individuals without phenotypes. The agreement between the output of the two softwares has been confirmed by comparing variance estimates obtained through both packages.

The mixed linear models (MLM) can be presented as:

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e}$$

Where $\boldsymbol{y}$ is the vector of observations, $\boldsymbol{\tau}$ is a vector of fixed effects, $\boldsymbol{X}$ is an incidence matrix referring the observations to their corresponding fixed effect levels, described further below, $\boldsymbol{u}$ is a vector of breeding values treated as random effects and $\boldsymbol{e}$ is a vector of residual effects, assumed to be normally distributed with parameters $N(0, \sigma_e^2 \boldsymbol{I})$, where σ_e^2 is the residual variance and $\boldsymbol{I}$ is identity matrix, and $\boldsymbol{Z}$ is an incidence matrix referring observations to their

corresponding random effects. Random effects $\boldsymbol{u}$ were assumed to be normally distributed with parameters $N(0, \sigma_a^2 \boldsymbol{A})$ for pedigree based analyses (PED), where σ_a^2 is the additive genetic variance and $\boldsymbol{A}$ is a numerator relationship matrix (Henderson, 1975). For genomic analyses, random effects were assumed to be distributed with parameters $N(0, \sigma_g^2 \boldsymbol{G})$, where σ_g^2 is the genetic variance explained by markers and **G** is a genomic relationships matrix. Each chip provided different source of information, hence **G** varied across chip densities. The **G** matrix was constructed in ACTA software package (Gray et al., 2012) which follows the second method of calculating the GRM as defined by VanRaden (2008), implemented in Yang et al. (2011). The genomic relationship matrix is defined as:

$$\mathbf{G}=\mathbf{W}\mathbf{W}^{\mathrm{T}}/\mathrm{N}$$

Where **W** is an incidence matrix of size *P* x *N* (where P is the number of individuals and N is the number of SNPs).

The elements of **W** matrix are centred and standardized as:

$$\boldsymbol{W}_{ij} = \frac{x_{ij} - 2p_i}{\sqrt{2p_i(1-p_i)}}$$

Where x_{ij} is the number of reference allele copies for i^{th} SNP for individual *j* and p_i is the frequency of that allele.

Thus, the elements of **G** (for individuals *i* and *k*) are:

$$\boldsymbol{G}_{jk} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1-p_i)}$$

where x_{ij} is the number of reference allele copies at i^{th} SNP for individual *j* and x_{ik} is the number for individual *k*.

The fixed effects fitted in the models are presented in Table 4.1. For reproductive traits, the effect of mate was included to account for possible variation caused by the sire. Most females were paired with the same male throughout the breeding period. There were 14 females which were paired with a different male in the late breeding period, for these females the fixed effect of a mate fitted to LFERT and LHOF was the replacement rooster.

Each of the mixed linear model analyses was iterated until convergence, i.e. until the difference in variance estimates and likelihood between iterations changed by less than 1% (for the estimates) and 0.002*I, where I is the current iteration number (for likelihood) (Gilmour et al., 2006).

3.2 Accuracy and Bias of Predictions

Accuracies of phenotype predictions (r_P) were calculated in GenStat (Payne et al. 2009) as a correlation between recorded phenotype and predicted BV. Both phenotype and (G)EBV were corrected for fixed effects. Accuracy of breeding value prediction (r_A) was obtained by dividing r_P by the square root of heritability. The estimate of the heritability used for this standardization was taken from pedigree based analyses on the same dataset. Standard error of the accuracy of phenotype prediction was calculated as $SE_P = 1/\sqrt{df}$ where *df* is the number of degrees of freedom left in the TST after accounting for fixed effect levels. The standard error of the breeding value prediction was obtained by scaling the SE_P by the square root of heritability.

The bias of prediction was calculated in GenStat from the regression of phenotypes on the predicted BVs, with fixed effects accounted for in the model, as described in Chapter 3.

4. Results

4.1 Variance estimation

The estimates of variance components obtained through REML analysis are presented in Table 4.3. Based on the reference estimates of REML using **A** matrix (PED), traits analysed were characterised by a wide range of heritability estimates, from low heritability of MORT at 0.04 (SE 0.02) to moderately high heritability of BWT at 0.35 (SE 0.04). The differences between the estimates of genetic variance for a given trait were mostly insignificant, however, estimates of genetic variance based on genomic relationship matrices were generally lower than the reference. The proportion of variance that was not identified using low density 2k chip varied between 13.6% (EFERT) to 54% (MORT) of the reference PED value. For most traits (BWT, HHP, EFERT, LFERT and EHOF) increasing marker density brought the genomic estimates closer to the reference. The exception to this was noted in variance estimates of MORT and LHOF, for which estimates of the genetic variance based on higher density chips exceeded the PED reference.

Table 4.3 Estimates of variance components for the 7 traits, obtained from pedigree based (PED) analysis and genomic analyses run on different chips. σ_A^2 - variance of the direct genetic effect of an individual, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance, h^2 - heritability. Standard errors are given in brackets (SE).

Trait	Method	σ_A^2 (SE)	σ_e^2 (SE)	σ_P^2 (SE)	h^2 (SE)
BWT					
	PED	55.62 (7.46)	104.14 (5.20)	159.76 (5.04)	0.35 (0.04)
	2k	41.84 (4.92)	116.74 (3.63)	158.57 (5.04)	0.26 (0.03)
	42k	47.90 (5.41)	110.42 (3.77)	158.32 (4.96)	0.30 (0.03)
	74k	50.14 (5.61)	109.62 (3.79)	159.76 (5.08)	0.31 (0.03)
	175k	45.24 (5.09)	109.95 (3.79)	155.18 (4.71)	0.29 (0.03)
	431k	52.13 (5.81)	109.14 (3.81)	161.28 (4.71)	0.32 (0.03)
HHP					
	PED	146.52 (28.92)	487.16 (24.27)	633.68 (20.74)	0.23 (0.04)
	2k	94.36 (17.02)	529.34 (18.41)	623.70 (19.67)	0.15 (0.03)
	42k	124.66 (21.10)	505.61 (19.28)	630.27 (20.44)	0.20 (0.03)
	74k	128.02 (21.78)	505.82 (19.40)	633.84 (20.75)	0.20 (0.03)
	175k	120.77 (20.21)	502.98 (19.37)	623.75 (19.82)	0.19 (0.03)
	431k	131.61 (22.54)	506.55 (19.51)	638.16 (21.13)	0.21 (0.03)
EFERT					
	PED	24.72 (7.61)	232.36 (9.34)	257.08 (8.15)	0.10 (0.03)
	2k	21.36 (5.59)	235.31 (8.30)	256.66 (8.13)	0.08 (0.02)
	42k	20.74 (5.81)	235.34 (8.50)	256.08 (8.08)	0.08 (0.02)
	74k	21.59 (6.11)	234.97 (8.55)	256.56 (8.13)	0.08 (0.02)
	175k	19.67 (5.63)	235.16 (8.56)	254.83 (7.99)	0.08 (0.02)
	431k	22.56 (6.40)	234.76 (8.57)	257.32 (8.20)	0.09 (0.02)
LFERT					
	PED	41.39 (11.80)	333.48 (13.87)	374.86 (12.11)	0.11 (0.03)
	2k	32.45 (8.07)	340.29 (12.13)	372.74 (11.97)	0.09 (0.02)
	42k	36.07 (8.99)	336.59 (12.44)	372.66 (11.99)	0.10 (0.02)
	74k	38.58 (9.54)	335.32 (12.50)	373.90 (12.10)	0.10 (0.02)
	175k	35.63 (8.81)	335.07 (12.51)	370.70 (11.85)	0.10 (0.02)
	431k	40.45 (9.96)	334.89 (12.53)	375.34 (12.23)	0.11 (0.03)
EHOF					
	PED	18.54 (4.07)	69.71 (3.57)	88.25 (3.00)	0.21 (0.04)
	2k	10.53 (2.19)	76.06 (2.75)	86.53 (2.80)	0.12 (0.02)
	42k	12.71 (2.55)	74.07 (2.84)	86.78 (2.84)	0.15 (0.03)
	74k	13.95 (2.73)	73.34 (2.85)	87.29 (2.89)	0.16 (0.03)
	175k	12.94 (2.53)	73.25 (2.86)	86.19 (2.79)	0.15 (0.03)
	431k	15.06 (2.90)	72.86 (2.86)	87.92 (2.95)	0.17 (0.03)
LHOF					
	PED	16.44 (5.32)	131.17 (5.93)	147.61 (4.95)	0.11 (0.04)
	2k	18.22 (4.02)	128.68 (4.94)	146.90 (4.97)	0.12 (0.03)
	42k	19.59 (4.39)	127.10 (5.10)	146.69 (4.96)	0.13 (0.03)
	74k	20.38 (4.59)	126.94 (5.13)	147.32 (5.02)	0.14 (0.03)
	175k	18.58 (4.23)	127.17 (5.15)	145.75 (4.87)	0.13 (0.03)
	431k	21.62 (4.85)	126.58 (5.16)	148.20 (5.11)	0.15 (0.03)
MORT					
	PED	37.97 (19.16)	896.86 (28.09)	934.82 (24.50)	0.04 (0.02)
	2k	17.46 (11.22)	914.90 (25.61)	932.36 (24.25)	0.02 (0.01)
	42k	46.83 (17.26)	889.04 (26.57)	935.87 (24.62)	0.05 (0.02)
	74k	51.92 (18.47)	877.66 (26.87)	941.84 (25.23)	0.07 (0.02)
	175k	53.93 (17.78)	880.33 (26.78)	934.26 (24.54)	0.06 (0.02)
	431k	64.17 (20.56)	877.66 (26.87)	941.84 (25.23)	0.07 (0.02)

For MORT the PED reference estimate was exceeded with chip density of 42k and almost doubled at the highest density of 431k. For LHOF, genomic estimates of genetic variance across all chip densities exceeded those obtained using pedigree only. The changes in the error variance reflected the changes in the estimates of genetic variance, i.e. error variance decreased with increasing marker density toward the reference value of pedigree based predictions.

4.2 BIAS OF PREDICTIONS

The regression coefficients from different methods and chips are presented in Table 4.4. Due to low numbers of individuals with records for mortality and fertility traits, bias of predictions for these traits was characterised by large standard errors, thus none of the differences observed between methodologies were statistically significant. The regression coefficient of BLUP PED predictions departed from expected unity, to a different degree in different traits. Lowest bias of predictions was found for HHP at 0.96 (SE 0.21), with EFERT following at 0.91 (0.35). For these two traits, the departure was not statistically significant. In contrast, MORT predictions grossly overestimated the range of actual BVs, with regression coefficient found at 0.13 (SE 0.69). The only trait which showed an underestimation of the BVs through BLUP methodology based on pedigree only was LHOF, at 1.67 (SE 0.61). Genomic predictions of most traits, with exception for EFERT and EHOF, also overestimated the range of true breeding values, as approximated by the phenotype. Bias of predictions using marker information was larger than BLUP PED bias for BWT, HHP and LFERT. For EFERT, although bias of predictions from low density chip (2k) was larger than bias of pedigree reference, predictions using higher density chips for this trait resulted in nearly unbiased estimates, with slight tendency to underestimate the range of true values. Regression coefficients larger than 1 were also found for EHOF, with a pattern of decreasing bias noted with increasing chip density. The regression coefficients of genomic predictions for MORT were closer to 1 than the BLUP PED prediction, however, despite the extreme difference between 0.13 found for BLUP PED and 0.85 found for predictions based on 2k markers could not be statistically proven due to large standard errors (t=0.66, p>0.05).

Table 4 Bias of (G)EBV predictions as regression coefficient for different chip densities, standard errors given in brackets.

Method	BWT	HHP	EFERT	LFERT	EHOF	LHOF	MORT
PED	0.71 (0.08)	0.96 (0.21)	0.91 (0.35)	0.70 (0.41)	0.76 (0.25)	1.67 (0.61)	0.13 (0.69)
2k	0.64 (0.07)	0.89 (0.17)	0.65 (0.33)	0.47 (0.40)	1.34 (0.25)	0.77 (0.34)	0.85 (0.85)
42k	0.69 (0.07)	0.81 (0.17)	1.05 (0.37)	0.49 (0.41)	1.20 (0.24)	0.84 (0.36)	0.53 (0.45)
74k	0.67 (0.07)	0.79 (0.18)	1.04 (0.37)	0.46 (0.41)	1.21 (0.24)	0.85 (0.36)	0.46 (0.45)
175k	0.78 (0.08)	0.80 (0.20)	1.13 (0.41)	0.55 (0.46)	1.33 (0.26)	0.85 (0.40)	0.48 (0.47)
431k	0.68 (0.07)	0.81 (0.18)	1.02 (0.38)	0.49 (0.41)	1.19 (0.24)	0.82 (0.36)	0.40 (0.41)

4.3 ACCURACY OF PREDICTIONS

Figure 4.1 shows the accuracy of BV predictions calculated as a correlation between residuals of the phenotype and (G)EBVs.

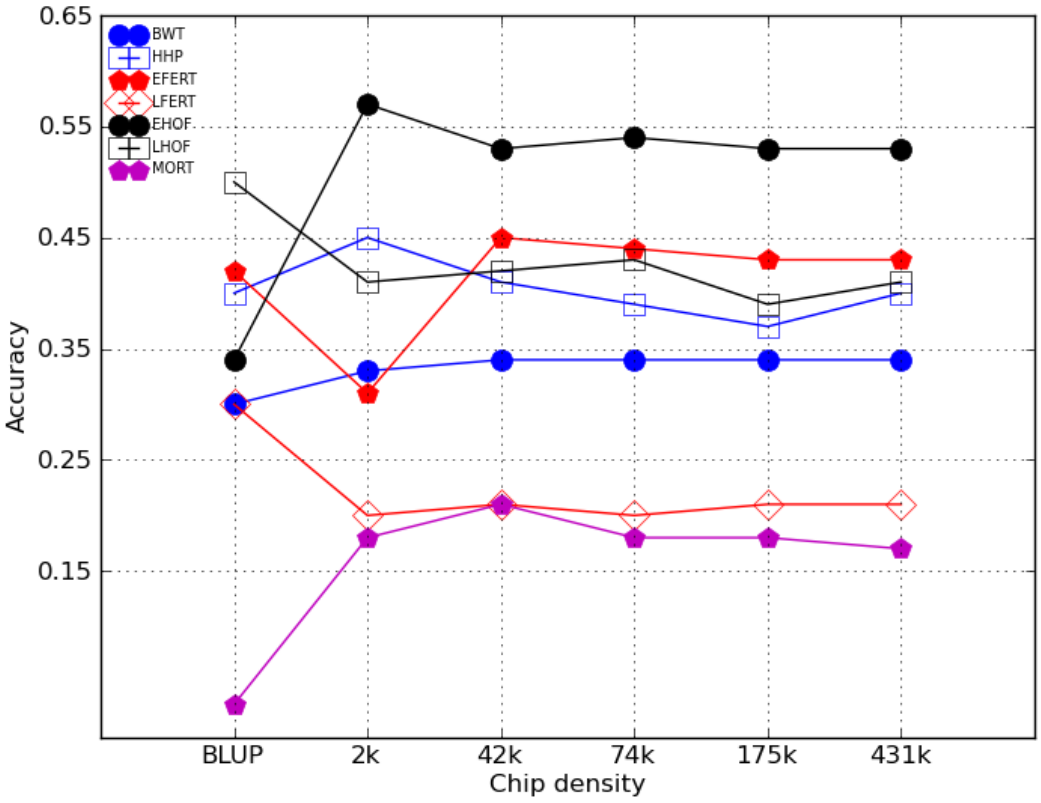

Figure 4.1 Accuracy of breeding value prediction, calculated as a correlation between residuals of phenotype and (G)EBVs, corrected for fixed effects and plotted against the method (BLUP PED vs GBLUP) and chip density.

Accuracy of genomic predictions exceeded the accuracy of predictions based on BLUP PED for most traits, particularly at high marker densities. For EHOF, the benefits of using marker

information increased the accuracy by 68%. For MORT, the accuracy of pedigree based predictions was not significantly different from zero, while genomic predictions for this trait reached an accuracy of 0.21 (SE 0.20) for 42k chip. Reproductive traits recorded late in life (LHOF and LFERT) showed no benefit on incorporating marker data in this dataset, with highest accuracies found for BLUP PED predictions. Except for EFERT, for which increasing marker density from 2k to 42k improved accuracy by 45%, high density panels did not bring additional benefits over low density chips.

The accuracy of phenotype (r_P) and breeding value (r_A) predictions are shown in Table 4.5. Due to low numbers of records available in TST, the differences observed between accuracies of predictions failed to reach statistical significance.

Table 4.5 Estimates of heritability and accuracy of phenotype (r_P) and breeding value (r_A) prediction for pedigree based prediction (PED) and genomic predictions at different marker densities, with standard errors given in brackets.

Trait	Parameter	PED	2k	42k	75k	175k	431k
BWT	h^2	0.35 (0.04)	0.26 (0.03)	0.30 (0.03)	0.31 (0.03)	0.29 (0.03)	0.32 (0.03)
	r_P	0.18 (0.02)	0.19 (0.02)	0.20 (0.02)	0.20 (0.02)	0.20 (0.02)	0.20 (0.02)
	r_A	0.30 (0.03)	0.33 (0.03)	0.34 (0.03)	0.34 (0.03)	0.34 (0.03)	0.34 (0.03)
HHP	h^2	0.23 (0.04)	0.15 (0.03)	0.20 (0.03)	0.20 (0.03)	0.19 (0.03)	0.21 (0.03)
	r_P	0.19 (0.04)	0.22 (0.04)	0.20 (0.04)	0.19 (0.04)	0.18 (0.04)	0.19 (0.04)
	r_A	0.40 (0.08)	0.45 (0.08)	0.41 (0.08)	0.39 (0.08)	0.37 (0.08)	0.40 (0.08)
EFERT	h^2	0.10 (0.03)	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)	0.09 (0.02)
	r_P	0.13 (0.05)	0.10 (0.05)	0.14 (0.05)	0.14 (0.05)	0.13 (0.05)	0.13 (0.05)
	r_A	0.42 (0.16)	0.31 (0.16)	0.45 (0.16)	0.44 (0.16)	0.43 (0.16)	0.43 (0.16)
LFERT	h^2	0.11 (0.03)	0.09 (0.02)	0.10 (0.02)	0.10 (0.02)	0.10 (0.02)	0.11 (0.03)
	r_P	0.10 (0.06)	0.07 (0.06)	0.07 (0.06)	0.07 (0.06)	0.07 (0.06)	0.07 (0.06)
	r_A	0.30 (0.18)	0.20 (0.18)	0.21 (0.18)	0.20 (0.18)	0.21 (0.18)	0.21 (0.18)
EHOF	h^2	0.21 (0.04)	0.12 (0.02)	0.15 (0.03)	0.16 (0.03)	0.15 (0.03)	0.17 (0.03)
	r_P	0.15 (0.05)	0.26 (0.05)	0.24 (0.05)	0.25 (0.05)	0.24 (0.05)	0.24 (0.05)
	r_A	0.34 (0.10)	0.57 (0.10)	0.53 (0.10)	0.54 (0.10)	0.53 (0.10)	0.53 (0.10)
LHOF	h^2	0.11 (0.04)	0.12 (0.03)	0.13 (0.03)	0.14 (0.03)	0.13 (0.03)	0.15 (0.03)
	r_P	0.17 (0.06)	0.14 (0.06)	0.14 (0.06)	0.14 (0.06)	0.13 (0.06)	0.14 (0.06)
	r_A	0.50 (0.18)	0.41 (0.18)	0.42 (0.18)	0.43 (0.18)	0.39 (0.18)	0.41 (0.18)
MORT	h^2	0.04 (0.02)	0.02 (0.01)	0.05 (0.02)	0.07 (0.02)	0.06 (0.02)	0.07 (0.02)
	r_P	0.01 (0.04)	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)	0.04 (0.04)	0.03 (0.04)
	r_A	0.03 (0.20)	0.18 (0.20)	0.21 (0.20)	0.18 (0.20)	0.18 (0.20)	0.17 (0.20)

5. DISCUSSION

The aim of this study was to evaluate the benefits of using genomic markers in the prediction of GEBVs for seven broiler traits of different genetic architecture and with different challenges regarding data recording. By fitting GBLUP models which utilized Genomic Relationship Matrices constructed using different marker densities, the effect of chip density on the estimations was evaluated. Although the differences between the accuracies found in this study failed to reach statistical significance due to low numbers of records, genomic methods improved on the accuracies found through pedigree-base analyses for most traits, particularly for MORT and EHOF. It is expected that with the increasing numbers of phenotyped and genotyped individuals the accuracy of genomic predictions will continue to increase.

The benefits of GBLUP over BLUP are expected to be largest for traits with lower heritability, as with high heritability traits the phenotypic information in BLUP is usually sufficient to provide an accurate estimate of the BV. This has been illustrated in poultry, where traits with moderately high heritability showed great response to selection, e.g. in body weight. The high heritability of BWT in this study can thus explain the minor improvement of GBLUP over BLUP PED found in the analysed dataset. However, as the genotyped population grows, the accuracy of GEBVs will continue to increase, as it has been proven to be dependent not only on the heritability of the trait, but also on the number of genotyped and phenotyped individuals in TRN (Daetwyler et al., 2008, Daetwyler et al., 2010b, Meuwissen et al., 2001). In contrast, accuracy of pedigree-based predictions is not likely to change. Increasing the number of collateral relatives (siblings, cousins) in pedigree based BLUP predictions improves the predictions, however, EBVs based on pedigree only do not account for the Mendelian sampling term. Thus, an EBV estimated using pedigree information only has an upper limit of accuracy of 0.7, when parental EBVs are estimated without error. In contrast, genomic prediction could potentially achieve an accuracy of 1, if the markers captured all the genetic variance and their effects were accurately estimated.

The same expectations of increasing the benefits of GBLUP over BLUP with increasing numbers of TRN animals will apply to HHP, which has the second largest heritability out of the traits analysed in this study. Accuracy of predictions for this trait showed some fluctuation, with the highest accuracy of GEBVs found for 2k chip predictions and exceeding the accuracy of BLUP PED, however the magnitude of these changes is small. Aside from the expected rise of the accuracy of prediction with an increased number of TRN individuals, as mentioned above, genomic prediction for this trait offers an additional benefit

in the form of accurate predictions of male GEBVs and calculation of GEBVs early in life of a selection candidate. In BLUP predictions for sex-limited traits such as HHP, or other fertility traits studied in this experiment, the EBVs of male selection candidates are based on phenotypic measurements of their female relatives. By utilizing marker information, an accurate estimate of the BV can be obtained for young selection candidates of both sexes, without the need for the trade-off between time, when highly accurate male EBVs are obtained upon collection of sister and daughter records, and accuracy, when the EBVs are calculated based on female ancestors only. The benefit of accurate estimation of male EBVs is particularly important considering that a selected male will mate several females.

The benefit of estimating male GEBVs through the use of the marker information applies also to the other reproductive traits analysed in this experiment. It is of particular interest for fertility, for which some ambiguity exists as to which of the parents contributes more to the observed trait. Furthermore, the analyses of this trait are also complicated by the age effects and permanent environment variance (Wolc et al., 2009). Wolc *et al*. (2009) showed that although the genetic variance for fertility is fairly constant over the breeding period, the effect of permanent environment increases with age of the breeders. This can partly explain the poorer performance of predictions for late fertility in this study, as models used for EFERT and LFERT predictions did not account for permanent environment. EFERT showed an increase of the prediction accuracy of GEBVs vs EBVs calculated using BLUP PED, which was expected, considering previous findings in the literature concerning predictions for traits with low heritability (Goddard and Hayes, 2007). The estimates of the heritability for fertility traits found in this dataset at 0.10 and 0.11 (SE 0.03) for EFERT and LFERT respectively exceed values previously reported, ranging from 0.06 (SE 0.01) for fertility estimated using a repeatability model (Sapp et al., 2004) and 0.07 for female fertility for one week of egg production, averaged over weeks, using a repeatability model (Wolc et al., 2009). Due to the lack of standard error estimates in other studies and different models used for calculation of the variance component estimates, direct comparison of these estimates is not possible.

Similarly to fertility, hatchability is a complex trait posing challenges to both recording procedures, and interpretation of the data. Recorded on adult and reproducing females, hatchability is a typical sex-limited trait, thus it is expected to benefit from genomic predictions, particularly for male selection candidates, as described above. From a biological perspective, interpretation of the records for this trait is difficult. Hatchability is known to be affected by multiple factors, both environmental and genetic, with the latter asserted by both

the dam and the chick itself. The environmental factors include diet, storage and incubation temperature and length, and are now routinely accounted for in the management of the breeding flocks (as reviewed by King'Ori, 2011). Factors with known genetic background commonly quoted in the literature relate mostly to egg quality traits, such as egg weight, shell thickness and porosity and overall shell quality (as reviewed by King'Ori, 2011). These traits have been extensively studied in layers (as reviewed by Dunn, 2011). They have been shown to be affected by multiple QTLs, for example there are over 30 QTLs identified for egg weight, 11 QTLs for egg shell weight, 18 for shell thickness, 22 for shell strength and 4 for egg shape (QTLdb). Those QTLs are spread among multiple chromosomes, including sex chromosome Z. The heritability of these traits has also been determined, e.g. the heritability of shell thickness has been estimated at 0.34 (SE 0.09) in a brown egg layer line (Zhang et al., 2005). The overall estimates of hatchability in broilers differ between studies and periods considered, with estimates ranging from 0.14 (as reviewed by Szwaczkowski, 2003) to 0.30 (Hunton, 1969). The estimates found in this study fall within this range, with heritability of late hatchability estimated lower than that of EHOF, at 0.11 and 0.21 respectively. The difference in heritability can partly explain the lower accuracy of LHOF predictions found in this study. Decreased heritability leads to increased sampling error of the estimates of marker effects, thus reducing accuracy of selection (Meuwissen et al., 2001). Alternatively, the marker set used in the analyses presented could include markers in linkage disequilibrium with QTLs affecting the EHOF and without an effect on LHOF. It has been shown previously that different QTLs influence early and late production periods (Abasht et al., 2009). A QTL of large effect on EHOF could explain the observed benefits of GBLUP over BLUP PED for this trait, which were not observed in LHOF.

Mortality is the last trait in the presented results. Though obviously very important for the production and welfare of the chickens, this trait is very difficult to evaluate and despite the breeders' best efforts, selection against mortality has not been very successful, with mortality rates remaining unchanged (Besbes and Ducrocq, 2003). Mortality rates in juvenile broilers are typically estimated at around 5% (Leitner et al., 1989, Gonzalez-Recio et al., 2008), with similar rates estimated for layer breeders (Besbes and Ducrocq, 2003). These low numbers of cases are inevitably accompanied by large standard errors of the estimates, thus leading to an inability to detect the genetic variance. As such, the estimates of heritability for this trait are frequently low, e.g. 0.02 - 0.03 found for chick mortality (Long et al., 2007). Mortality is a compound trait, which encompasses multiple traits, from disease resistance, to metabolic traits. There have been few studies looking into causative QTLs affecting overall mortality, with 2 QTLs identified for early mortality, 1 QTL for late and 1 for total mortality (QTLdb).

However, there are multiple studies looking into more precise causes of disease and death in broilers. The proportion of each particular mortality cause is likely to change over the lifetime of a bird. The mortality in young chicks is in most cases believed to be related to infectious diseases (Leitner et al., 1989), thus early mortality can be treated as an indicator of immune responses. The number of QTLs identified for disease susceptibility traits of chickens are constantly growing, with over 50 QTLs already mapped to 21 chromosomes. In contrast, mortality of adult and semi-adult birds is frequently linked to metabolic disorders, such as ascites, which is positively correlated to body weight and feed efficiency (Pakdel et al., 2005). The numbers of QTLs related to such traits are also spread over multiple chromosomes. Considering such large numbers of contributing factors, it follows that informative analysis of MORT will require large numbers of records. Furthermore, due to the positive correlation between some of those metabolic disorders and body mass and feed efficiency, selection against the causes of mortality might be counteracted by intensive selection for production traits. Models which included marker information in the presented analysis of adult broiler mortality identified a larger genetic variance component than pedigree based analyses, with the estimate obtained from 431k chip being nearly twice as big as the pedigree reference, however, the large standard errors of these estimates did not allow any firm conclusions. This observation, although opposite to common expectations of the proportion of variance explained by markers being lower than the total genetic variance, has been reported for this trait before (Gonzalez-Recio et al., 2008). Thus, it could be speculated that the marker set contained SNPs in linkage with QTLs related to traits affecting survival. Further analyses of larger numbers of records will hopefully provide more information on this finding.

Considering low heritability of MORT and its positive correlation to production traits, the lack of progress in selection against this trait using traditional, pedigree-based methods is hardly surprising. The results of this study confirm previous findings that genomic prediction has the potential to improve on the accuracy of selection for this trait (Gonzalez-Recio et al., 2008); however, further studies are necessary to confirm significance of this improvement. It is also possible that the analysis of the MORT would bring more informative results after data transformation allowing for the binary nature of the records.

The results of this study showed an interesting pattern, where the accuracy of predictions did not align with the expectations based on the heritability of the traits. It has been widely accepted that traits with higher heritability are expected to have higher accuracies than traits with lower heritability, due to smaller sampling error of haplotype effects (Meuwissen et al.,

2001). In the presented results, BWT which had the highest heritability of the analysed traits (h^2=0.35) had moderate accuracy of prediction (r=0.34, SE 0.04), which was exceeded by accuracy of traits with lower heritabilities, e.g. EHOF, with h^2=0.21 and r=0.53 (SE 0.10). Considering the large standard error of the estimates, increasing the numbers of individuals with phenotypes will be crucial for more precise evaluation of the performance of GP.

The estimates of bias for most of the traits (all except LHOF) analysed through BLUP PED indicate that the predicted EBVs overestimate the range of true breeding values. For BWT, the bias was relatively large, which can be explained by selection pressure applied to this trait and not included in the analysed data. Bias of genomic predictions for this trait was similar to the one obtained using BLUP PED, with the differences observed not reaching statistical significance. For other traits, comparison of bias estimates across methods and chip densities is not a reliable source of information as the standard errors far exceeded the magnitude of the differences between compared estimates.

With exception of MORT and LHOF, markers failed to capture all the genetic variance identified by the reference of BLUP PED. This observation is not surprising, and has been previously observed and examined in human studies of diseases (Yang et al., 2010, Zuk et al., 2012). Yang *et al*. (2010) suggested that missing variance can be explained by imperfect linkage between SNPs and causative loci, and the magnitude of the effect and allele frequency of the latter. The authors speculated that the imperfect linkage will be particularly detrimental when the trait in question is subjected to selection, which results in reduced Minor Allele Frequency (MAF) at the causative locus. All of the traits used in the presented experiment belong to the range of breeding goals used in routine evaluations of the commercial stock. As such, they have been subject to selection, with varying intensity. Thus, it follows that many of the causative loci for these traits are likely to have low MAF. However, the marker screening criteria used in this study are not as stringent as those presented in Yang *et al.*, (2010), with MAF threshold values set at 0.01, as opposed to 0.1 in the published paper, therefore the proportion of the variance lost due to imperfect linkage is likely to account for only a part of the observed loss. Considering that markers used in this study are spread over 27 chromosomes, out of 38 available in the chicken genome, it is plausible that a part of the variance missed in the current dataset is determined by loci on the unrepresented chromosomes, i.e. microchromosomes 16, and 28-38. Although their physical size is small, the microchromosomes of the chicken genome are characterized by high gene density (Hillier et al., 2004), thus it is plausible that the genetic variance undetected by available marker set could be at least partially explained by QTLs on these chromosomes.

However, this hypothesis would require further testing. At the time of the study, a QTL search for loci affecting body weight at 35 days shows 27 QTLs located on 11 chromosomes (QTLdb), and these chromosomes were included in the set analysed. Microchromosomes are currently under-represented in most association studies, due to their specific qualities which make their cloning difficult (Groenen et al., 2009). Alternatively, another possible explanation for the missing variance is the lack of sufficient numbers of phenotypic records used for the estimation of marker effects.

6. CONCLUSIONS

This preliminary study into the use of GP in broiler chickens showed a great promise, with the accuracy of GEBVs estimated between 0.18 (SE 0.20) for MORT and 0.57 (SE 0.10) for EHOF. For most traits, the accuracy of GP exceeded the accuracy of BLUP PED, with the benefits of marker based methods expected to increase with growing numbers of genotyped individuals. The effect of marker density on the accuracy of predictions was limited, particularly when marker numbers exceeded 42k. Although further studies are required to confirm the expected benefits of GS, the accuracies achieved in this study show a great potential for improvement of the progress in sex-limited and mortality traits in broilers.

7. REFERENCES

ABASHT, B., SANDFORD, E., ARANGO, J., SETTAR, P., FULTON, J. E., O'SULLIVAN, N. P., HASSEN, A., HABIER, D., FERNANDO, R. L., DEKKERS, J. C. & LAMONT, S. J. 2009. Extent and consistency of linkage disequilibrium and identification of DNA markers for production and egg quality traits in commercial layer chicken populations. *BMC Genomics,* 10 Suppl 2**,** S2.

ABDOLLAHI-ARPANAHI, R., NEJATI-JAVAREMI, A., PAKDEL, A., MORADI-SHAHRBABAK, M., MOROTA, G., VALENTE, B., KRANIS, A., ROSA, G. & GIANOLA, D. 2014. Effect of allele frequencies, effect sizes and number of markers on prediction of quantitative traits in chickens. *Journal of Animal Breeding and Genetics*, 131, 123 - 133

BESBES, B. & DUCROCQ, V. 2003. Use of mixed model methodology in breeding strategies for layers. *Poultry Breeding and Biotechnology. W. Muir, and S. Aggrey, ed. CABI Publishing, Wallingford, UK***,** 127-146.

CALUS, M. P. L., MEUWISSEN, T. H. E., DE ROOS, A. P. W. & VEERKAMP, R. F. 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics,* 178**,** 553-561.

CHEN, C. Y., MISZTAL, I., AGUILAR, I., TSURUTA, S., MEUWISSEN, T. H. E., AGGREY, S. E., WING, T. & MUIR, W. M. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. *J Anim Sci,* 89**,** 23-28.

CLEVELAND, M. A., HICKEY, J. M. & FORNI, S. 2012. A common dataset for genomic analysis of livestock populations. *G3: Genes| Genomes| Genetics,* 2**,** 429-435.
DAETWYLER, H. D., HICKEY, J. M., HENSHALL, J. M., DOMINIK, S., GREDLER, B., VAN DER WERF, J. H. J. & HAYES, B. J. 2010a. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. *Animal Production Science,* 50**,** 1004-1010.
DAETWYLER, H. D., PONG-WONG, R., VILLANUEVA, B. & WOOLLIAMS, J. A. 2010b. The impact of genetic architecture on genome-wide evaluation methods. *Genetics,* 185**,** 1021-31.
DAETWYLER, H. D., VILLANUEVA, B. & WOOLLIAMS, J. A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One,* 3**,** e3395.
DE GREEF, K. H., JANSS, L. L., VEREIJKEN, A. L., PIT, R. & GERRITSEN, C. L. 2001. Disease-induced variability of genetic correlations: ascites in broilers as a case study. *Journal of Animal Science,* 79**,** 1723-33.
DE LOS CAMPOS, G., HICKEY, J. M., PONG-WONG, R., DAETWYLER, H. D. & CALUS, M. P. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics,* 193**,** 327-45.
DUNN, I. 2011. Poultry breeding for egg quality: traditional and modern genetic approaches. Improving the safety and quality of eggs and egg products. Volume 1: Egg chemistry, production and consumption pp 245-260.
FOOTE, R. H. 2003. Fertility estimation: a review of past experience and future prospects. *Animal Reproduction Science,* 75**,** 119-139.
FOSSUM, O., JANSSON, D., ETTERLIN, P. & VAGSHOLM, I. 2009. Causes of mortality in laying hens in different housing systems in 2001 to 2004. *Acta Veterinaria Scandinavica,* 51**,** 3.
GILMOUR, A. R., GOGEL, B. J., CULLIS, B. R., WELHAM, S. J. & THOMPSON, R. 2006. *{ASREML} User Guide Release 2.0*.
GODDARD, M. E. & HAYES, B. 2007. Genomic Selection. *Journal of Animal Breeding and Genetics,* 124**,** 323 - 330.
GONZALEZ-RECIO, O., GIANOLA, D., LONG, N., WEIGEL, K. A., ROSA, G. J. M. & AVENDANO, S. 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. *Genetics,* 178**,** 2305-2313.
GRAY, A., STEWART, I. & TENESA, A. 2012. Advanced complex trait analysis. *Bioinformatics,* 28**,** 3134-6.
GROENEN, M., MEGENS, H.-J., ZARE, Y., WARREN, W., HILLIER, L., CROOIJMANS, R., VEREIJKEN, A., OKIMOTO, R., MUIR, W. & CHENG, H. 2011. The development and characterization of a 60K SNP chip for chicken. *BMC Genomics,* 12**,** 274.
GROENEN, M. A. M., WAHLBERG, P., FOGLIO, M., CHENG, H. H., MEGENS, H. J., CROOIJMANS, R. P. M. A., BESNIER, F., LATHROP, M., MUIR, W. M., WONG, K., GUT, I. & ANDERSSON, L. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research,* 19**,** 510-519.
HENDERSON, C. R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,* 31**,** 423-447.
HICKEY, J., KINGHORN, B., TIER, B., WILSON, J., DUNSTAN, N. & VAN DER WERF, J. 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. *Genetics Selection Evolution,* 43**,** 12.
HILLIER, L. W., MILLER, W., BIRNEY, E., WARREN, W., HARDISON, R. C., PONTING, C. P., BORK, P., BURT, D. W., GROENEN, M. A. M., DELANY, M. E., DODGSON, J. B., CHINWALLA, A. T., CLIFTEN, P. F., CLIFTON, S. W.,

DELEHAUNTY, K. D., FRONICK, C., FULTON, R. S., GRAVES, T. A., KREMITZKI, C., LAYMAN, D., MAGRINI, V., MCPHERSON, J. D., MINER, T. L., MINX, P., NASH, W. E., NHAN, M. N., NELSON, J. O., ODDY, L. G., POHL, C. S., RANDALL-MAHER, J., SMITH, S. M., WALLIS, J. W., YANG, S. P., ROMANOV, M. N., RONDELLI, C. M., PATON, B., SMITH, J., MORRICE, D., DANIELS, L., TEMPEST, H. G., ROBERTSON, L., MASABANDA, J. S., GRIFFIN, D. K., VIGNAL, A., FILLON, V., JACOBBSON, L., KERJE, S., ANDERSSON, L., CROOIJMANS, R. P. M., AERTS, J., VAN DER POEL, J. J., ELLEGREN, H., CALDWELL, R. B., HUBBARD, S. J., GRAFHAM, D. V., KIERZEK, A. M., MCLAREN, S. R., OVERTON, I. M., ARAKAWA, H., BEATTIE, K. J., BEZZUBOV, Y., BOARDMAN, P. E., BONFIELD, J. K., CRONING, M. D. R., DAVIES, R. M., FRANCIS, M. D., HUMPHRAY, S. J., SCOTT, C. E., TAYLOR, R. G., TICKLE, C., BROWN, W. R. A., ROGERS, J., BUERSTEDDE, J. M., WILSON, S. A., STUBBS, L., OVCHARENKO, I., GORDON, L., LUCAS, S., MILLER, M. M., INOKO, H., SHIINA, T., KAUFMAN, J., SALOMONSEN, J., SKJOEDT, K., WONG, G. K. S., WANG, J., LIU, B., WANG, J., YU, J., YANG, H. M., NEFEDOV, M., KORIABINE, M., DEJONG, P. J., GOODSTADT, L., WEBBER, C., DICKENS, N. J., LETUNIC, I., SUYAMA, M., TORRENTS, D., VON MERING, C., et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature,* 432**,** 695-716.

HUNTON, P. 1969. Variance and covariance of hatchability and some of its components in the chicken 1. *British Poultry Science,* 10**,** 261-272.

KING'ORI, A. 2011. Review of the factors that influence egg fertility and hatchability in poultry. *International Journal of Poultry Science,* 10**,** 483-493.

KRANIS, A., GHEYAS, A., BOSCHIERO, C., TURNER, F., YU, L., SMITH, S., TALBOT, R., PIRANI, A., BREW, F., KAISER, P., HOCKING, P., FIFE, M., SALMON, N., FULTON, J., STROM, T., HABERER, G., WEIGEND, S., PREISINGER, R., GHOLAMI, M., QANBARI, S., SIMIANER, H., WATSON, K., WOOLLIAMS, J. & BURT, D. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics,* 14**,** 59.

LEITNER, G., DAN HELLER, E. & FRIEDMAN, A. 1989. Sex-related differences in immune response and survival rate of broiler chickens. *Veterinary Immunology and Immunopathology,* 21**,** 249-260.

LONG, N., GIANOLA, D., ROSA, G., WEIGEL, K. & AVENDANO, S. 2007. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. *Journal of Animal Breeding and Genetics,* 124**,** 377-389.

MEUWISSEN, T. Genomic Selection: The Future of Marker Assisted Selection and Animal Breeding. FAO International Workshop "Marker Assisted Selection: A Fast Track to Increase in Genetic Gain in Plant and Animal Breeding?"q, 2003 Turin.

MEUWISSEN, T. & GODDARD, M. 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution,* 33**,** 605 - 634.

MEUWISSEN, T., HAYES, B. & GODDARD, M. E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics,* 157**,** 1819 - 1829.

MUIR, W. M., WONG, G. K., ZHANG, Y., WANG, J., GROENEN, M. A. M., CROOIJMANS, R. P. M. A., MEGENS, H. J., ZHANG, H. M., MCKAY, J. C., MCLEOD, S., OKIMOTO, R., FULTON, J. E., SETTAR, P., O'SULLIVAN, N. P., VEREIJKEN, A., JUNGERIUS-RATTINK, A., ALBERS, G. A. A., LAWLEY, C. T., DELANY, M. E. & CHENG, H. H. 2008. Review of the initial validation and characterization of a 3K chicken SNP array. *Worlds Poultry Science Journal,* 64**,** 219-225.

PAKDEL, A., VAN ARENDONK, J. A. M., VEREIJKEN, A. L. J. & BOVENHUIS, H. 2005. Genetic parameters of ascites-related traits in broilers: correlations with feed efficiency and carcase traits. *British Poultry Science,* 46**,** 43 - 53.

PREISINGER, R. 2012. Genome-wide selection in poultry. *Animal Production Science,* 52**,** 121-125.

PRYCE, J. E. & DAETWYLER, H. D. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Animal Production Science,* 52**,** 107-114.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. W., DALY, M. J. & SHAM, P. C. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics,* 81**,** 559-575.

QTLDB, C. *Chicken QTLdb* [Online]. Available: http://www.animalgenome.org/cgi-bin/QTLdb/GG/summary [Accessed 16/06/2014.

SAPP, R., REKAYA, R., MISZTAL, I. & WING, T. 2004. Male and female fertility and hatchability in chickens: A longitudinal mixed model approach. *Poultry Science,* 83**,** 1253-1259.

SZWACZKOWSKI, T. 2003. Use of mixed model methodology in poultry breeding: estimation of genetic parameters. *In:* MUIR, W. M. & AGGREY, S. E. (eds.) *Poultry Genetics, Breeding and Biotechnology.* CABI Publishing.

VANRADEN, P. M. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science,* 91**,** 4414-4423.

WOLC, A. & OLORI, V. Genetics of hatchability-egg quality from the perspective of a chick. Proc. 6th Eur. Poult. Genet. Symp, Będlewo, Poland, 2009.

WOLC, A., WHITE, I. M., OLORI, V. E. & HILL, W. G. 2009. Inheritance of fertility in broiler chickens. *Genetics Selection Evolution,* 41**,** 47.

WONG, G. K., LIU, B., WANG, J., ZHANG, Y., YANG, X., ZHANG, Z., MENG, Q., ZHOU, J., LI, D., ZHANG, J., NI, P., LI, S., RAN, L., LI, H., LI, R., ZHENG, H., LIN, W., LI, G., WANG, X., ZHAO, W., LI, J., YE, C., DAI, M., RUAN, J., ZHOU, Y., LI, Y., HE, X., HUANG, X., TONG, W., CHEN, J., YE, J., CHEN, C., WEI, N., DONG, L., LAN, F., SUN, Y., YANG, Z., YU, Y., HUANG, Y., HE, D., XI, Y., WEI, D., QI, Q., LI, W., SHI, J., WANG, M., XIE, F., ZHANG, X., WANG, P., ZHAO, Y., LI, N., YANG, N., DONG, W., HU, S., ZENG, C., ZHENG, W., HAO, B., HILLIER, L. W., YANG, S. P., WARREN, W. C., WILSON, R. K., BRANDSTROM, M., ELLEGREN, H., CROOIJMANS, R. P., VAN DER POEL, J. J., BOVENHUIS, H., GROENEN, M. A., OVCHARENKO, I., GORDON, L., STUBBS, L., LUCAS, S., GLAVINA, T., AERTS, A., KAISER, P., ROTHWELL, L., YOUNG, J. R., ROGERS, S., WALKER, B. A., VAN HATEREN, A., KAUFMAN, J., BUMSTEAD, N., LAMONT, S. J., ZHOU, H., HOCKING, P. M., MORRICE, D., DE KONING, D. J., LAW, A., BARTLEY, N., BURT, D. W., HUNT, H., CHENG, H. H., GUNNARSSON, U., WAHLBERG, P., ANDERSSON, L., KINDLUND, E., TAMMI, M. T., ANDERSSON, B., WEBBER, C., PONTING, C. P., OVERTON, I. M., et al. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature,* 432**,** 717-22.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. & VISSCHER, P. M. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet,* 42**,** 565-569.

ZHANG, L. C., NING, Z. H., XU, G. Y., HOU, Z. C. & YANG, N. 2005. Heritabilities and genetic and phenotypic correlations of egg quality traits in brown-egg dwarf layers. *Poultry Science,* 84**,** 1209 - 1213.

ZUK, O., HECHTER, E., SUNYAEV, S. R. & LANDER, E. S. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences,* 109**,** 1193-1198.

CHAPTER 5

SOURCES OF INFORMATION USED IN GENOMIC PREDICTION

1. INTRODUCTION

Thanks to recent advances in genomic technologies, increasing amounts of genotypes are being produced worldwide for many livestock species. One of their primary uses in animal breeding is calculation of genome wide estimated breeding values (GEBVs). These estimates are expected to be more accurate than estimates obtained using traditional methods, utilizing phenotype and pedigree information only (Meuwissen et al., 2001).

The traditional approach to genetic improvement is based on BLUP methodology (Henderson, 1975), in which the breeding values of individuals are calculated using a numerator relationship matrix **A** scaled by the additive genetic variance. In the classical approach to BLUP, coefficients of the **A** matrix are twice the coefficients of coancestry, with the latter quantifying the probability that alleles of two gametes taken at random, one from each individual, are identical by descent - IBD (Falconer and Mackay, 1996). The estimates of the genetic variance obtained using the pedigree and phenotypes in studied population are relative to the base population, i.e. the first generation in the pedigree. Individuals in this population are assumed to be unrelated, non-inbred and originate from random breeding (Falconer and Mackay, 1996). The genetic resemblance between individuals further down the pedigree is quantified by relationship coefficients. In the pedigree-based approach, those coefficients take on values expected for given classes of relationships. This framework has been developed when no direct information could be obtained on the actual genes shared between individuals and allowed impressive genetic improvement in many livestock populations. With the advent of molecular data it has been shown that it can be improved upon through estimation of the realized relationships, obtained from observed genotypes (Hayes et al., 2009).

The use of genotypes allows a more accurate estimation of relationships between individuals, as it identifies differences between relatives, e.g. full sibs, that arise as a result of Mendelian sampling of the parental gametes (Hayes et al., 2009). The expected values of the relationship coefficients for given pedigree relationships are therefore replaced by values

derived from the observed allele sharing between individuals (Goddard and Hayes, 2007). In practice, this is done by replacing **A** with the genomic relationship matrix, **G**, in an approach referred to as genomic prediction (GP) (Meuwissen et al., 2001). While the original approach to GP relied on explaining the genetic variance through utilizing Linkage Disequilibrium (LD) between markers and causative QTLs, the LD methodology can be also used to define relationships rather than estimate genotype at QTL (Hayes et al., 2009). This method could theoretically be used without the pedigree, as it traces relationships that precede those contained in the pedigree (Luan et al., 2012), encompassing covariances between individuals in the base generation through alleles that are identical by state (IBS). However, the accuracy of the estimates using an LD approach can deteriorate rather quickly over relatively few generations, as linkage is broken by recombination occurring during meiosis (Hartl and Clark, 1997). As the methodology behind the LD approach is based on the direct link between markers and phenotypes, the choice of SNPs used as markers and their location is expected to influence the results and efficacy of this approach.

However, irrespective of their effect on the trait, markers can provide invaluable information on the inheritance of chromosome segments, tracked from the base population down the pedigree, and forming an identity by descent (IBD) matrix (Meuwissen and Goddard, 2010). A Linkage Analysis (LA) approach combines the theoretical assumptions of the **A** matrix for the base population with observed allele sharing among genotyped individuals. As the genomic relationship matrix constructed using the LA approach ($\mathbf{G_{LA}}$) relies upon the family structure, genetic variants that appear in the base population are assumed to be distinct, despite being identical by state, which follows the classical assumptions of the base population consisting of unrelated individuals. The LA approach follows the classical method also in referring the assumed variance to the base population, i.e. the first generation in the pedigree. As the method utilizes markers to keep track of recombinations that occur in the genome, rather than to explain the phenotypic resemblance using markers, the LA approach is not expected to be affected by the choice of SNPs in the same degree as the LD approach.

From the assumptions of these three methods, it follows that the relationship sub-matrix of individuals in the first generations of the pedigree will be the same identity matrix in classical and LA approaches while in the LD approach it will carry values estimated directly from the data. As each of these methods represents a different source of information for use in estimation of animal's breeding values, a flexible approach to combining their contributions could provide an optimal use of genotypes and pedigree. Therefore, the

objective of this study was to evaluate A, LD and LA matrices and their mixtures when fitted to a BLUP model for estimation of Breeding Values (BVs) of broiler chickens. Fit of the model to data was assessed by the likelihood values, while the efficacy of predicting BVs of selection candidates was evaluated using accuracy and bias estimates. To check for the possible effect of SNP choice on the performance of the tested methods, matrices were calculated on two different in-silico chips.

2. MATERIALS

2.1 POPULATION

The population used for the following analyses has been already studied in Chapter 3, where its detailed description can be found. The phenotypes used were juvenile body weight, recorded at 35 days of age on both sexes (BWT) and available for all 5,416 individuals. The population was split into a training population (TRN) and testing population (TST), consisting of 3,146 and 2,270 individuals respectively. Phenotypes of TST individuals were masked for variance estimation and prediction of breeding values and were used to check accuracy and bias of predictions. The TST individuals were offspring and siblings of individuals in TRN and none of them had offspring with records included in TRN. The first individuals genotyped formed the base population of this study and were assumed unrelated and non-inbred. For individuals where only one parent was known, that parent was removed. As a result, the base population consisted of 288 individuals with no ancestors known.

2.2 SNP CHOICE

Genotypes used in the analyses consisted of 431K markers which passed quality control (QC) performed in Plink (Purcell et al., 2007) on the full range of SNP markers from Affymetrix Axiom 600K chip (Kranis et al., 2013). Table 5.1 gives the statistics of the markers failed due to particular screening criteria.

Table 5.1 Quality control statistics with the percentage of markers failed as a proportion of the original 625,995 SNPs. Some rejected markers failed more than one quality criteria.

Category	Quality criteria	Proportion of SNPs failed
Hardy Weinberg equilibrium	p<=0.001	3.6%
Missing	>0.05	5.0%
Minor Allele Frequency	<0.01	25.0%
Remaining		431,249 (69% of all)

While the original 600K genotypes included 28 chromosomes and several un-located contigs, only SNPs with known chromosome location were used. Furthermore, none of SNPs on chromosome 16 passed the QC. Thus, out of 38 chromosomes present in the chicken genome, the dataset included all macrochromosomes (chromosomes 1-5) and intermediate chromosomes (6 to 10), plus 17 out of 28 microchromosomes. Markers from sex chromosome Z were not included in the analyses.

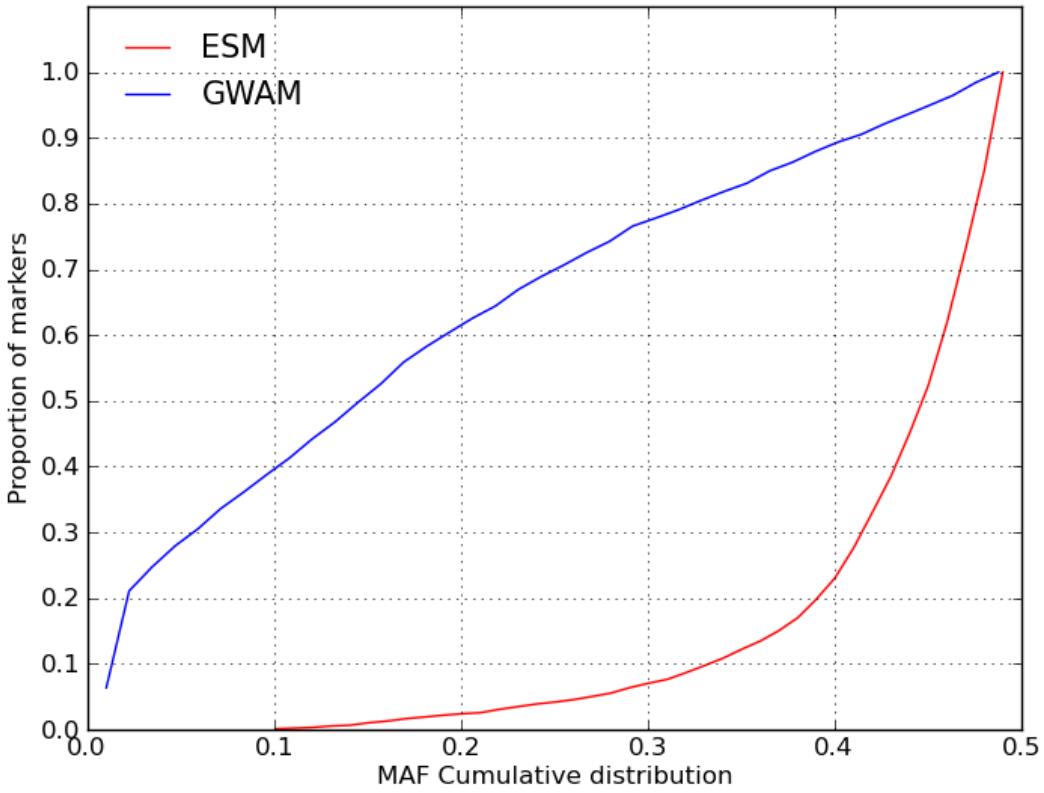

Figure 5.1 Cumulative distribution of Minor Allele Frequencies for the two chips

From the 431K SNPs that passed the QC, two in-silico chips were created with the same density of 27K SNPs (1k/chromosome). Markers for these chips were selected through:

- their effect on the trait assessed using GWAS - genome wide association study (genome wide associated markers - GWAM)

- location in the genome (evenly spaced markers - ESM).

The two chips differed in the average MAF of SNPs selected, as shown on Figure 5.1, with GWAM chip favouring SNPs with lower MAF and ESM chip favouring SNPs with higher MAF. The distances between markers differed between chips as well, as shown in Figure 5.2.

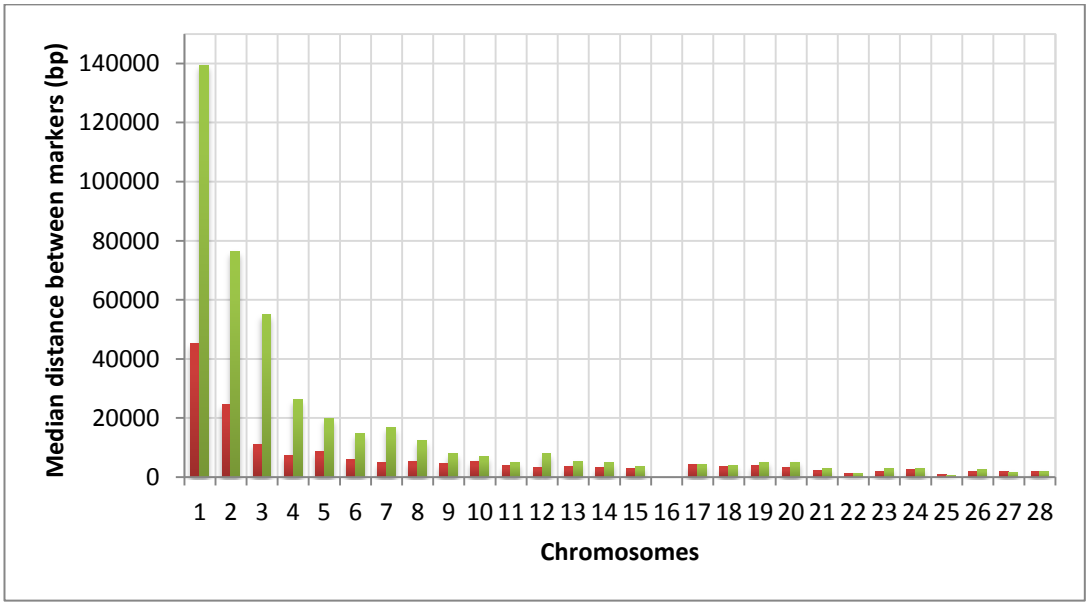

Figure 5.2 Median distance in base pairs between markers for GWAM (red bars) and ESM chips (green bars).

2.2.1 GWAM chip

In this approach, 27k SNPs were selected based on GWAS analysis of BWT, carried out in Plink (Purcell et al., 2007). The analysis used genotypes and phenotypes of individuals in the TRN only. SNPs were split between chromosomes and ranked according to their p-value. The top 1000 SNPs on each of the 27 chromosomes were selected. These included some that did not exceed the genome-wide significance threshold of GWAS hits.

2.2.2 ESM Chip

In the second approach, SNPs were selected according to their spacing on the linkage map rather than on the physical map. This accounted for variable recombination rates on different chromosomes, as the distances were calculated in centimorgans (cM). The linkage map was custom built by A. Kranis from Aviagen data (personal communication). As with the GWAM chip, a constant density of 1,000 markers on each of the chromosomes was selected. Maintaining the same number of SNPs per chromosome despite large differences in chromosome size in chickens inevitably led to a higher density of markers on the microchromosomes (11-28). In the final choice of SNPs for these equidistant panels, markers with high MAF were preferred. For chromosomes 25, 26, 27 and 28 which are classified as

microchromosomes and have much shorter sequences, the number of SNPs selected was 888, 998, 998 and 991 respectively.

3. METHODS

3.1 CALCULATION OF RELATIONSHIP MATRICES

The reference base for this study was a pedigree based numerator relationship matrix **A** (Henderson, 1975), calculated in the ASReml software package (Gilmour et al., 2006).

3.1.1 Linkage Analysis Relationship Matrix (G_{LA})

The calculation of the first of the genomic relationship matrices used followed Fernando and Grossman (1989), in which the markers are traced through generations up to the base population as defined by the pedigree. Thus, the part of $\mathbf{G_{LA}}$ relating to the base is an identity matrix, just as in **A**. The relationships in subsequent generations are improved upon as compared to pedigree, due to the use of marker information. The $\mathbf{G_{LA}}$ was calculated using LDMIP - Linkage Disequilibrium Multilocus Iterative Peeling method (Meuwissen and Goddard, 2010), which calculates relationships between parental gametes of two individuals, which results in 4 relationship estimates for each pair of individuals at every locus. The four gametic relationships are summed and divided by two to obtain a relationship at a locus. To obtain the total relationship between individuals, the relationships at all loci were subsequently averaged.

3.1.2 Linkage Disequilibrium Relationship Matrix (G_{LD})

The second genomic relationship matrix considers relationships both within and beyond the pedigree, allowing for relationships in the base population which are not documented in the pedigree. Thus, the relationship coefficients of the first individuals in the pedigree differ from those in **A** and $\mathbf{G_{LA}}$ matrices. The $\mathbf{G_{LD}}$ matrix was constructed in the ACTA software package (Gray et al., 2012) which follows methodology defined by Yang et al. (2011), described in detail in Chapter 3.

3.2 BLENDING MATRICES

As each of the above matrices utilizes a different source of information and bears different relationship coefficients, blending of such information may maximize the benefit of using marker information (Goddard et al., 2011). Mixing of the relationships was done by weighting the particular relationship coefficients according to:

$$\mathbf{M}=\lambda\mathbf{M}_1+(1-\lambda)\mathbf{M}_2$$

Two types of blending were considered: LDA, where $\mathbf{M}_1=\mathbf{G}_{\mathbf{LD}}$ and $\mathbf{M}_2=\mathbf{A}$ matrices, and LDLA where $\mathbf{M}_1=\mathbf{G}_{\mathbf{LD}}$ and $\mathbf{M}_2=\mathbf{G}_{\mathbf{LA}}$. The optimum blending was searched for, with λ moving from 0 to 1, at 0.1 steps. Options with λ=1 represented information obtained solely from the $\mathbf{G}_{\mathbf{LD}}$ matrix, while λ=0 sourced all information from the pedigree based matrix (either **A** in LDA or $\mathbf{G}_{\mathbf{LA}}$ in LDLA).

3.3 (G)EBV PREDICTION

All of the above-mentioned matrices were fitted as random effects in mixed linear models (MLM) in the ASReml package (Gilmour et al., 2006). Aside from the genetic effect, records were also corrected for fixed effects of hatch week and sex. The MLM is traditionally presented as:

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e}$$

Where $\boldsymbol{y}$ is the vector of observations, $\boldsymbol{X}$ is an incidence matrix associating the observations to their fixed effect levels, $\boldsymbol{\tau}$ is a vector of fixed effects and $\boldsymbol{Z}$ is an incidence matrix assigning observations to random effects. Vector $\boldsymbol{u}$ is a vector of breeding values treated as random genetic effects, assumed to be normally distributed with parameters $N(0, \sigma_A^2 \boldsymbol{M})$, where σ_A^2 is the additive genetic variance and $\boldsymbol{M}$ is a relationship matrix used in the given model (**A,** $\mathbf{G}_{\mathbf{LD}}$, $\mathbf{G}_{\mathbf{LA}}$ or their blended product). Vector $\boldsymbol{e}$ is a vector of random residual effects, assumed to be normally distributed with parameters $N(0,\sigma_e^2\boldsymbol{I})$, where σ_e^2 is the residual variance and $\boldsymbol{I}$ is identity matrix.

Each of the mixed linear model analyses was iterated until convergence, where the difference in LogLikelihood changed by less than 0.002 x *I*, where *I* is the current convergence number, and the variance estimates changed by less than 1% (Gilmour et al., 2006). The resulting logLikelihood estimates were later used to compare the fit of the models to data through the LogLikelihood Ratio Test (LRT). A null hypothesis for the comparison of the matrices was that λ=x, while the alternative hypothesis was that λ≠x. A statistically significant difference between lambda values was assumed when twice the difference in their log likelihoods was larger than 3.84. The confidence interval includes all the lambda values which were not rejected in LRT.

3.4 ACCURACY AND BIAS

While the logLikelihood estimates describe how well the models fit the data, practical application of genomic prediction depends on accuracy and bias of the predicted BVs. The accuracy was estimated in the same way as in Chapters 2 and 3, i.e. the accuracy of phenotype prediction (r_P) was calculated as a correlation between the (G)EBVs and the phenotypes of TST individuals after correcting both for fixed effects of hatch week and sex (i.e. correlation was calculated between the residuals), and accuracy of BV prediction (r_A) was obtained by dividing the r_p by the $\sqrt{h^2}$. A constant h^2 value of 0.35 was used for this purpose, which is the estimate of heritability obtained using pedigree based models run on this data, and stands in agreement with previous reports on the heritability of juvenile body weight in broilers (Gaya et al., 2006, de Verdal et al., 2011). The standard error of the correlations was calculated as $SE_P = 1/\sqrt{df}$ where *df* is the number of degrees of freedom left in the TST after accounting for fixed effect. The fixed effect of hatch week within the TST sample had 10 levels, while sex had two levels. Thus, the SE_P=0.02. The standard error of the breeding value prediction (SE_A) was obtained by scaling the SE_P by the square root of heritability, and resulted in SE_A=0.03.

Bias of the predictions was calculated from a regression of the phenotypes on the predicted (G)EBVs, with fixed effects accounted for in the regression. The regressions were fitted in GenStat software (Payne et al., 2009).

4. RESULTS

Although the two types of chips used for creating the relationship matrices between individuals consisted of nearly the same number of SNPs, the choice of the markers changed the profile of the likelihood for models fitted to the data. To facilitate the presentation of the results, findings from analyses carried out with each chip will be presented separately, with a comparison between the same blending method applied to two chips covered at the end of this section.

4.1 GWAM CHIP

4.1.1 Likelihood of the Models (GWAM chip)

Figure 5.3 presents the likelihood profile of LDA and LDLA analyses based on the GWAM chip. For both LDA and LDLA the likelihood profile was similar, differing primarily for small λ lambda range, where LDLA showed improved likelihood over LDA. The likelihood of the model incorporating $\mathbf{G_{LA}}$ matrix only was higher than that incorporating matrix **A**

alone, at 924.48 and 911.03 respectively. The change in likelihood values between consecutive λ values was large at low lambda (i.e. $0.0 < \lambda < 0.3$), but flattened out for values above 0.4. Maximum likelihood was attained at 0.7 for both LDA and LDLA, with support interval including lambda values between 0.5 and 0.8 (inclusive).

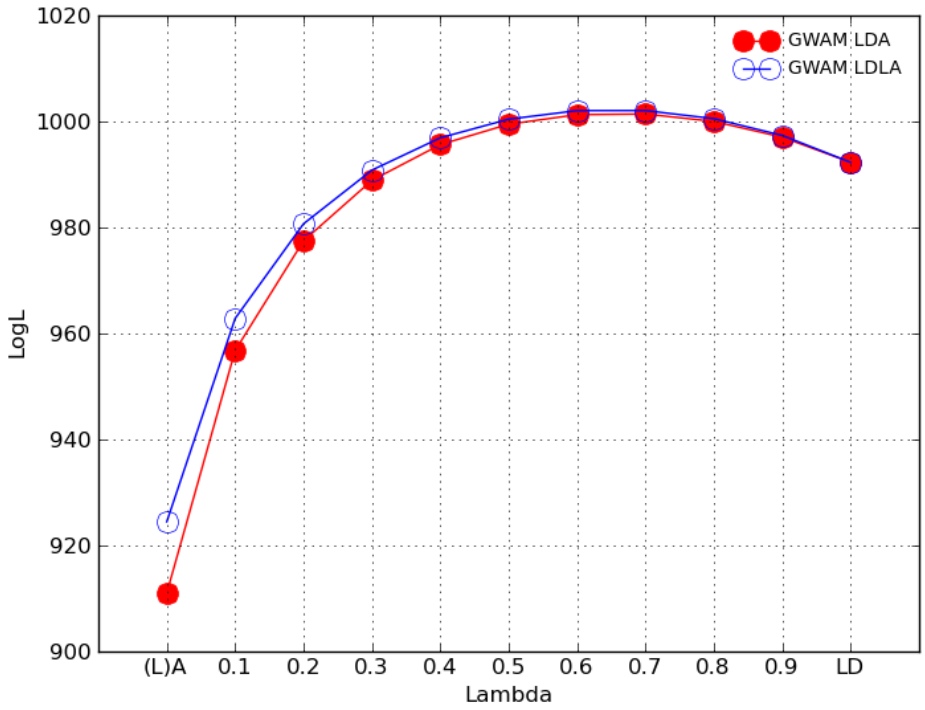

Figure 5.3 LogLikelihood Profile of LDA and LDLA models run using markers from the GWAM chip, with different lambda values.

Table 5.2 presents the estimates of the variance components obtained through REML run on the data using markers from GWAM chip and using both LDA and LDLA type of blending. The estimates of the genetic variance (σ^2_A) obtained using the original matrices (i.e. without blending) were highest when $\mathbf{G_{LA}}$ was fitted (LDLA at λ=0.0), however they were not significantly different from the estimates obtained from fitting the **A** matrix only (LDA at λ=0.0). The estimates of the σ^2_A obtained by fitting $\mathbf{G_{LD}}$ only (i.e. λ=1) were considerably lower than those of both $\mathbf{G_{LA}}$ and **A**, however, given large standard errors they were not significantly different either.

For both LDA and LDLA, there was a consistent pattern in the change of the variance estimates with increasing values of λ. The estimates of the heritability (h^2) obtained in LDLA for λ>0 were somewhat lower than in equivalent LDA models, however, the differences

were mostly small. For σ^2_A and h^2, initial addition of LD information (i.e. increasing λ from 0.0 to 0.1) brought an increase in these estimates to their highest value (i.e. heritability of 0.43 for LDA and 0.4 for LDLA), which was later gradually reduced to reach the lowest value when only LD information was used (i.e. h^2= 0.28 at λ=1.0). This pattern was reversed for error variance (V_E), which initially dropped at λ=0.1 and then gradually increased to highest value at λ=1.0. Most of the differences between estimates were not statistically significant, with exception of the differences between highest and lowest heritability estimates in LDA.

Table 5.2 Estimates of heritability (h^2) and variance components from LDA and LDLA analyses run using markers from the GWAM chip. σ^2_A - variance of the direct genetic effect of an individual, σ^2_e - residual variance, σ^2_P - total phenotypic variance. Standard errors are given in brackets.

λ	LDA			LDLA		
	σ^2_A	σ^2_e	h^2	σ^2_A	σ^2_e	h^2
0.0	54.94 (7.41)	104.54 (5.21)	0.35 (0.04)	57.26 (7.08)	103.78 (4.74)	0.36 (0.04)
0.1	68.43 (7.80)	92.25 (5.05)	0.43 (0.04)	64.25 (7.10)	96.25 (4.53)	0.40 (0.04)
0.2	66.53 (7.44)	92.06 (4.79)	0.42 (0.04)	62.13 (6.79)	96.13 (4.36)	0.39 (0.03)
0.3	63.53 (7.12)	93.59 (4.57)	0.40 (0.04)	59.61 (6.55)	97.14 (4.22)	0.38 (0.03)
0.4	60.60 (6.83)	95.62 (4.38)	0.39 (0.04)	57.31 (6.35)	98.54 (4.11)	0.37 (0.03)
0.5	57.84 (6.58)	97.88 (4.20)	0.37 (0.03)	55.23 (6.19)	100.15 (4.00)	0.36 (0.03)
0.6	55.23 (6.34)	100.27 (4.05)	0.36 (0.03)	53.28 (6.05)	101.93 (3.90)	0.34 (0.03)
0.7	52.66 (6.11)	102.78 (3.90)	0.34 (0.03)	51.34 (5.91)	103.89 (3.81)	0.33 (0.03)
0.8	49.99 (5.88)	105.46 (3.77)	0.32 (0.03)	49.22 (5.76)	106.09 (3.72)	0.32 (0.03)
0.9	46.94 (5.62)	108.39 (3.66)	0.30 (0.03)	46.64 (5.57)	108.64 (3.64)	0.30 (0.03)
1.0 (LD)	43.08 (5.28)	111.73 (3.57)	0.28 (0.03)	43.08 (5.28)	111.73 (3.57)	0.28 (0.03)

4.1.2 Accuracy and Bias of Predictions (GWAM chip)

Figure 5.4 presents accuracies of BV prediction obtained from LDA and LDLA using the GWAM chip.

Given the size of the training population, the best accuracy for this chip was achieved when marker information was supported by pedigree structure, i.e. using $\mathbf{G_{LA}}$, at 0.34 (SE 0.03). Accuracy of pedigree based prediction was lower, at 0.30 (SE 0.03). The lowest accuracy of predictions was found for predictions based on $\mathbf{G_{LD}}$ matrix. Although predictions based on LD approach were least accurate, mixing this source of information with pedigree based relationship at λ_{LDA}=0.1 improved upon the accuracy obtained through fitting **A** matrix alone. For both LDA and LDLA, the trend in changes of accuracy with increasing λ was almost linear, particularly for λ between 0.2 and 0.8.

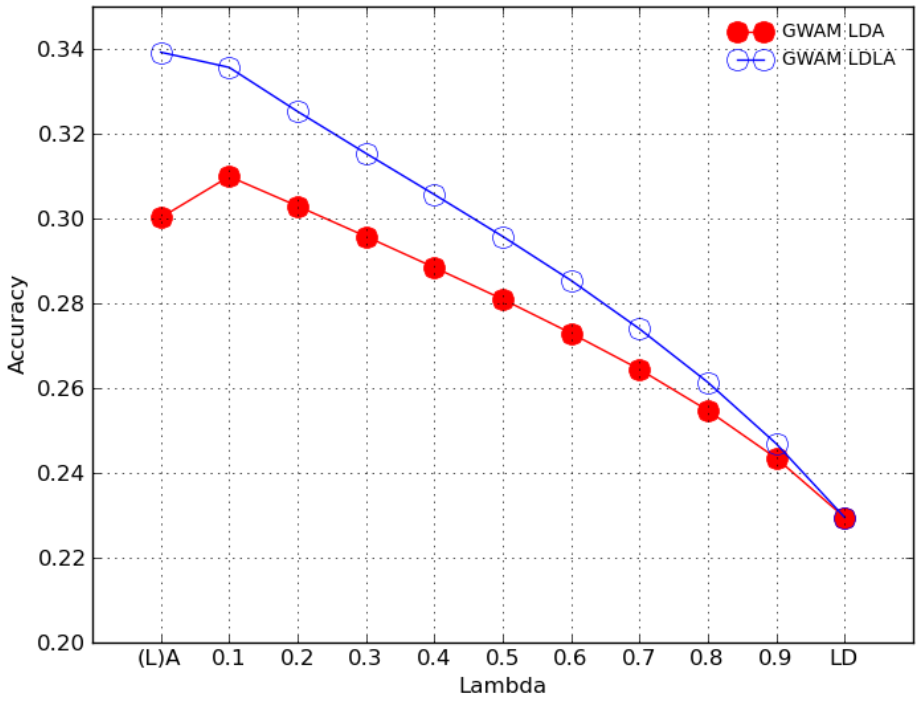

Figure 5.4 Accuracy of LDA and LDLA (G)EBV predictions based on the GWAM chip.

Table 5.3 shows the regression coefficients obtained from regressing actual phenotypes of the TST population on their predicted BVs, obtained from LDA and LDLA analyses, using both chips.

For GWAM, the least bias was found for predictions using $\mathbf{G}_{\mathbf{LA}}$ matrix only, at 0.77 (SE 0.08). Similarly for LDA, the maximum value of the regression coefficient was observed at λ=0, i.e. using **A**. For both methods, fitting pedigree based matrices was significantly better than regression of $\mathbf{G}_{\mathbf{LD}}$ based BVs. For all lambda values, mixing with $\mathbf{G}_{\mathbf{LA}}$ matrix gave less bias than using **A** matrix. However, the differences between **A** and $\mathbf{G}_{\mathbf{LA}}$ at the same λ values were not statistically significant.

Table 5.3 Bias of the (G)EBV predictions obtained from LDA and LDLA approaches, based on GWAM and ESM chips. Standard errors are given in brackets.

λ	GWAM		ESM	
	LDA	LDLA	LDA	LDLA
0.0	0.71 (0.08)	0.77 (0.08)	0.71 (0.08)	0.73 (0.08)
0.1	0.62 (0.07)	0.67 (0.07)	0.71 (0.08)	0.72 (0.08)
0.2	0.58 (0.07)	0.63 (0.07)	0.71 (0.08)	0.71 (0.08)
0.3	0.56 (0.07)	0.60 (0.07)	0.70 (0.07)	0.71 (0.08)
0.4	0.54 (0.07)	0.58 (0.07)	0.70 (0.07)	0.70 (0.07)
0.5	0.52 (0.07)	0.56 (0.07)	0.69 (0.07)	0.69 (0.07)
0.6	0.51 (0.07)	0.53 (0.07)	0.68 (0.07)	0.69 (0.07)
0.7	0.49 (0.07)	0.51 (0.07)	0.68 (0.07)	0.68 (0.07)
0.8	0.47 (0.07)	0.49 (0.07)	0.68 (0.07)	0.68 (0.07)
0.9	0.46 (0.07)	0.46 (0.07)	0.68 (0.07)	0.68 (0.07)
1.0 (LD)	0.44 (0.07)	0.44 (0.07)	0.68 (0.08)	0.68 (0.07)

4.2 ESM CHIP

4.2.1 Likelihood of the Models (ESM chip)

Figure 5.5 presents the likelihood profile of LDA and LDLA analyses based on ESM chip, with different λ values.

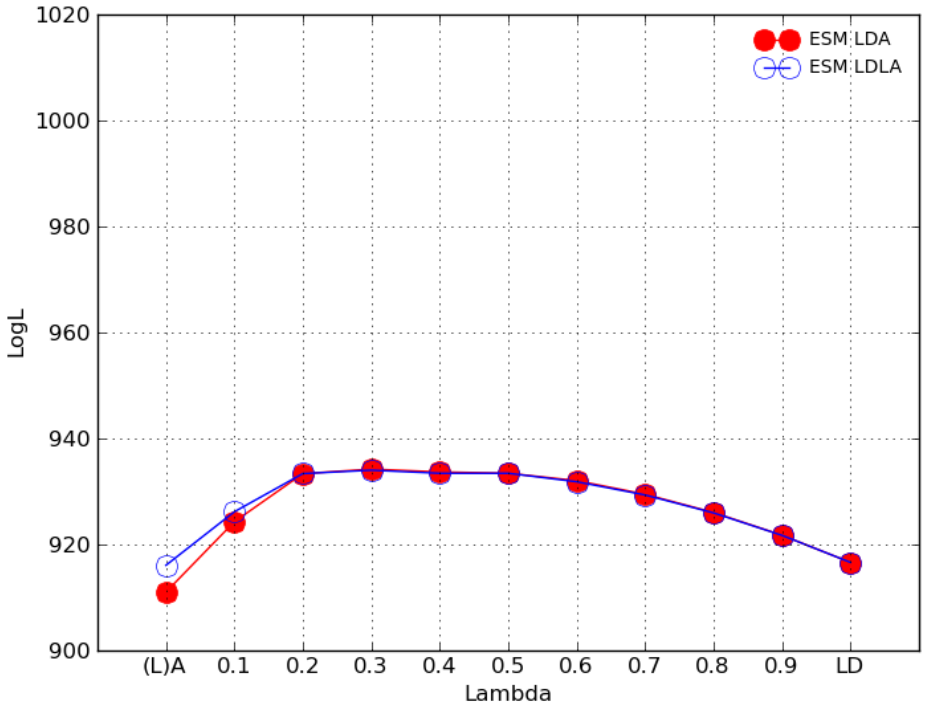

Figure 5.5 LogLikelihood Profile of LDA and LDLA models run using markers from the ESM chip, with different lambda values.

The likelihood profiles of LDA and LDLA based on ESM chip were similar, with a relatively flat, approximately bell-shaped distribution. The likelihood increased gradually

from starting values at lambda 0 to the maximum attained at λ=0.3, and then steadily decreased to the likelihood for $\mathbf{G_{LD}}$ at λ=1.0. The likelihood of the pedigree based analysis (i.e. using **A**) was the lowest of all recorded at 911.03 and significantly lower than any of the models including genomic information. The likelihood value at λ=0.0 for the model incorporating $\mathbf{G_{LA}}$ matrix only was higher than that incorporating **A** matrix alone, at 916.08. Likelihood of the model fitting $\mathbf{G_{LD}}$ was very similar to the one using $\mathbf{G_{LA}}$, at 916.62, thus making the profile of LDLA models more symmetrical than that of LDA. The 95% supporting interval for the highest likelihood was found between lambda 0.2 and 0.5 for both LDA and LDLA.

Table 5.4 presents the estimates of the variance components obtained through REML run on the data using markers from the ESM chip with variable mixture proportions in LDA and LDLA. Profile of the changes for the genetic variance and heritability estimates was similar between methods, with initial peak at λ=0.1 followed by gradual decrease of the values towards the lowest estimates at λ=1.0. The estimates of the phenotypic variance were relatively conserved for both methods and at all lambda values, ranging between 159.5 at λ=0.0 and 161.8 at λ=0.8 for LDA, and 159.8 at λ=0.0 and 161.7 at λ=0.8 for LDLA. While the pattern of changes was the same between methods, differences between LDLA estimates at different lambda values were smaller in magnitude than those of LDA.

The estimate of the heritability was highest for λ=0.1 in both types of analyses, at 0.38 (SE 0.04) for LDA and 0.35 (SE 0.04) for LDLA. The estimate of the genetic variance obtained through fitting the **A** matrix was higher than the estimates obtained by fitting either $\mathbf{G_{LA}}$ or $\mathbf{G_{LD}}$, however the $\mathbf{G_{LA}}$ estimate was much closer to the pedigree reference than $\mathbf{G_{LD}}$. The estimates of the heritability were not significantly different.

Table 5.4 Estimates of heritability (h^2) and variance components from LDA and LDLA analyses run using markers from the ESM chip. σ_A^2 - variance of the direct genetic effect of an individual, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance. Standard errors are given in brackets.

λ	LDA			LDLA		
	σ_A^2 (SE)	σ_e^2 (SE)	h^2 (SE)	σ_A^2 (SE)	σ_e^2 (SE)	h^2 (SE)
0.0	54.94 (7.41)	104.54 (5.20)	0.35 (0.04)	53.49 (6.77)	106.28 (4.72)	0.33 (0.04)
0.1	61.02 (7.61)	99.72 (5.13)	0.38 (0.04)	56.58 (6.81)	103.60 (4.63)	0.35 (0.04)
0.2	61.02 (7.41)	99.69 (4.94)	0.38 (0.04)	56.59 (6.70)	103.52 (4.52)	0.35 (0.03)
0.3	59.72 (7.19)	100.99 (4.75)	0.37 (0.04)	55.85 (6.59)	104.32 (4.42)	0.35 (0.03)
0.4	57.98 (6.98)	102.89 (4.57)	0.36 (0.04)	54.79 (6.49)	105.61 (4.31)	0.34 (0.03)
0.5	55.99 (6.77)	105.14 (4.40)	0.35 (0.03)	53.51 (6.39)	107.24 (4.21)	0.33 (0.03)
0.6	53.80 (6.56)	107.64 (4.24)	0.33 (0.03)	51.98 (6.27)	109.17 (4.11)	0.32 (0.03)
0.7	51.33 (6.33)	110.36 (4.09)	0.32 (0.03)	50.11 (6.13)	111.38 (4.01)	0.31 (0.03)
0.8	48.50 (6.06)	113.31 (3.96)	0.30 (0.03)	47.79 (5.95)	113.89 (3.92)	0.30 (0.03)
0.9	45.18 (5.75)	116.47 (3.84)	0.28 (0.03)	44.88 (5.70)	116.72 (3.83)	0.28 (0.03)
1.0 (LD)	41.24 (5.36)	119.84 (3.75)	0.26 (0.03)	41.24 (5.36)	119.84 (3.75)	0.26 (0.03)

4.2.2 Accuracy and Bias of the Predictions (ESM chip)

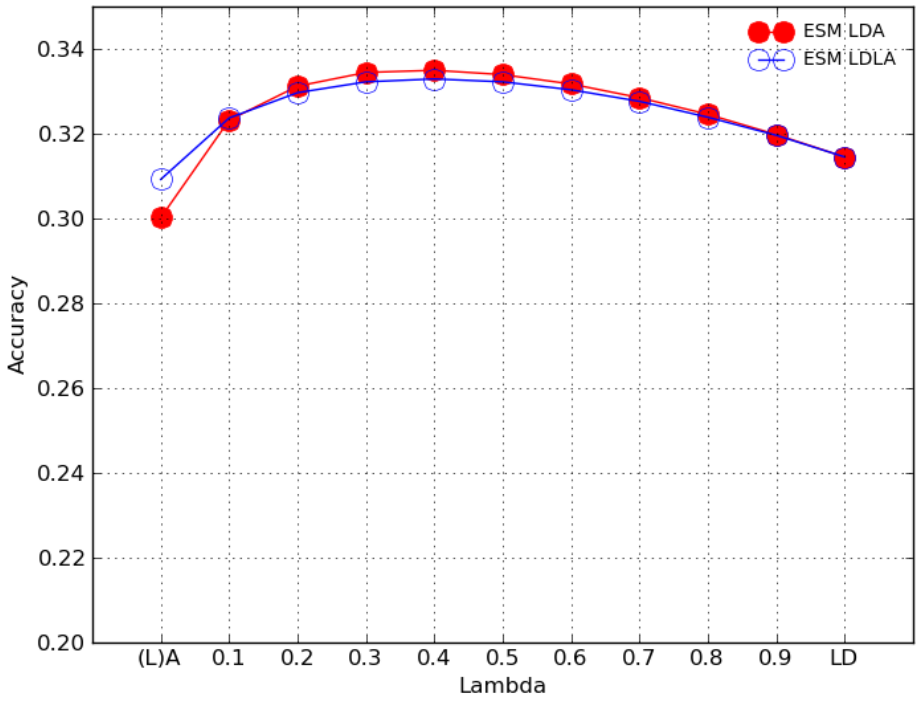

Figure 5.6 Accuracy of LDA and LDLA (G)EBV predictions based on the ESM chip.

Figure 5.6 presents accuracies of BV prediction obtained from LDA and LDLA using evenly spaced markers. The presented accuracies were obtained as the correlation between residuals after correcting the phenotypes and predicted BVs for fixed effects.

For both LDA and LDLA, the best accuracy, at 0.33 (SE 0.03), was achieved by using a combination of pedigree-based information (both **A** and $\mathbf{G_{LA}}$) with information sourced through LD approach, at λ=0.4. The benefit of including this type of information was particularly visible for LDA, which gave higher accuracy than LDLA for λ between 0.2 and 0.6. Overall, the changes observed were small, with the accuracy of predictions based on original matrices (i.e. without blending) similar, at 0.30 (SE 0.03) using pedigree only and 0.31 (SE 0.03) when genomic matrices were used.

Prediction bias of LDA and LDLA methods based on ESM chip was presented in Table 5.3. It can be seen that there was very little difference between using **A** and $\mathbf{G_{LA}}$ matrices in blending, and between different values of lambda used, as the estimates of bias were not significantly different between analyses.

4.3 COMPARISON OF CHIPS

The choice of markers had a considerable influence on the results observed. Selection of SNPs through GWAS hits resulted in much higher likelihood of $\mathbf{G_{LD}}$ based analyses than likelihood of ESM $\mathbf{G_{LD}}$ analyses, at 992.3 and 916.6 for λ=1.0. LA analyses were less affected by the choice of SNPs, with likelihood of LA analysis calculated at 924.48 and 916.08, for GWAM and ESM chips respectively. As a result, the profile of the likelihood changes with increasing lambda values was much flatter for the evenly spaced chip. The confidence interval for the maximum likelihood value contained different values for the two chips, including lambda values between 0.2 and 0.5 for ESM chip, and values between 0.5 and 0.8 for GWAM.

By using evenly spaced markers the likelihoods of genomic methods, i.e. using $\mathbf{G_{LD}}$ and $\mathbf{G_{LA}}$, became almost the same.

Estimates of the genetic variance based on an LD matrix using the ESM chip were slightly lower than respective results using the GWAM chip, while the estimates of error and total variance were higher. Thus, the estimates of heritability were higher for the GWAM chip, although most of the differences fell within the standard error range.

While the likelihood results indicate that the GWAM chip can model the variance better in the training population, accuracy and bias indicate that it is not as good in predicting

breeding values as models including evenly spaced markers. For the GWAM chip, the accuracy increased with a diminishing proportion of information sourced from the $\mathbf{G_{LD}}$ matrix to achieve highest value when all information was obtained from either **A** or $\mathbf{G_{LA}}$ matrices. Meanwhile, predictions based on evenly spaced markers achieved the highest accuracy when different sources of information were combined, however, the differences between methods were less obvious than in GWAM predictions. Predictions based on $\mathbf{G_{LD}}$ constructed on markers from the GWAM chip carried significantly larger bias than those using either pedigree based methods, both **A** and $\mathbf{G_{LA}}$, or using the same methodology on the ESM chip.

5. DISCUSSION

Real data collected on commercial broiler population was used to estimate the effect of different sources of information on the likelihood of linear models and accuracy of predictions in selection experiments. The optimum likelihood and accuracy were found for models which combined different sources of information, with best lambda estimated empirically on real data at 0.3 - 0.7. These values differ from published theoretical approximations, which for the same number of marker as used in this study approached 1 (VanRaden, 2008, Goddard et al., 2011). Using two in-silico chips with equal number but different selection of SNPs revealed a large effect of marker choice on the performance of methods used.

5.1 PERFORMANCE OF **A**, $\mathbf{G_{LA}}$ AND $\mathbf{G_{LD}}$

Matrices constructed on the GWAM chip showed a considerable difference in performance. Out of the three non-mixed matrices (**A,** $\mathbf{G_{LA}}$ and $\mathbf{G_{LD}}$), $\mathbf{G_{LD}}$ showed highest likelihood but was also least accurate in predicting GEBVs, with values predicted carrying a considerable bias. Increased likelihood of $\mathbf{G_{LD}}$ was expected, as GWAS analysis on basis of which SNPs were selected for this chip was carried out on the same training population as the variance component estimation. Reduced accuracy could be possibly caused by $\mathbf{G_{LD}}$ assigning high value to some SNPs that in fact had no effect on the trait, but exhibited specific associations within this particular TRN sample. While using these SNPs would result in higher likelihood of linear models fitted to the data, it would also explain the decrease in the prediction accuracy, as the chance links caused by population structure in TRN could be changed or missing in the TST population.

Inferiority of LD models compared to LA and A approaches observed on this chip can also be interpreted in terms of the type of information it was able to extract from marker data. Although the primary interpretation of the LD approach was based on linkage between markers and QTLs ensured by dense marker maps (Meuwissen et al., 2001), it has also been found that even in the absence of QTL associations the $\mathbf{G_{LD}}$ matrix can ensure non-zero accuracies by identifying family structures through the realized relationship matrix (Habier et al., 2007, Luan et al., 2012). This additional source of information depends on the architecture of the trait, with traits influenced by large QTLs being less affected by family connections, sourcing the majority of information from QTL associations (Luan et al., 2012). Body weight in chickens is a trait considered to be largely polygenic, with variance explained by multiple QTLs and their epistatic interactions (Carlborg et al., 2003). Thus, it would be expected that the accuracy of predictions using $\mathbf{G_{LD}}$ will be based on the benefits of using the realized relationship matrix and will therefore be equal to, or exceed the accuracy of pedigree based methods. This was not observed in analyses using $\mathbf{G_{LD}}$ based on GWAM. Significantly lowered accuracy of this method suggests that spurious QTL association over-ruled information sourced through the realized relationship matrix. These associations could be either an artefact without effect on the trait in the current population, or the linkage between QTLs and markers available was not strong enough for accurate predictions. Thus in the presented results the LD approach was not able to fully utilize familial information.

Supporting the GWAM chip with pedigree information in the form of $\mathbf{G_{LA}}$ exceeded the likelihood of the model based on **A**, and gave the best accuracy of predictions out of the models run using this chip. As this approach concentrates on the latest generations, it is more likely to detect segments of DNA affecting trait variance in current population than historical associations used in the LD approach. Compared to other studies utilizing the LA approach (Villanueva et al., 2005, Luan et al., 2012, Meuwissen et al., 2011), with majority analysing simulated populations of dairy cattle with test candidates being bulls evaluated on the basis of their daughter records, our study is unique in the abundance of the types of relationships presented in the pedigree, with reduced number of parent-offspring pairs and increased number of other relationships (average 16 chicks per sire and 5 chicks per dam in our data versus 45 daughters per bull in Luan et al. 2012). It is thus possible that benefits of using the LA approach in the analysed data were brought about by the unique pedigree structure. While $\mathbf{G_{LA}}$ provided the best accuracies in this dataset, possible future routine use of this method should be weighed against the costs of genotyping. The decay of LA information in

the absence of continued genotyping is more rapid than that of LD information (Habier et al., 2007).

In contrast to varied performance of matrices constructed on the GWAM chip, the performance of $\mathbf{G_{LD}}$ and $\mathbf{G_{LA}}$ constructed on the ESM chip was similar, both in terms of likelihood and accuracy. Similarity of the accuracy attained using $\mathbf{G_{LD}}$ and $\mathbf{G_{LA}}$ indicates that both these methods sourced most of the information from realized relationships, rather than through associations with QTLs. Lack of marked improvement through the use of $\mathbf{G_{LD}}$ indicates also low input from ancient covariances, i.e. preceding the formal pedigree. This result is supported by previous findings, suggesting that the effect of historical (i.e. beyond pedigree) covariances on covariances in more recent generations is limited (Luan et al., 2012, Nejati-Javaremi et al., 1997).

The benefits of using unmixed genomic relationship matrices in this study indicate a great potential for GS in broilers. Using just over 3,500 chickens in TRN allowed more accurate predictions than can be obtained using traditional methods. Pedigree based predictions of individuals without phenotypes are based mainly on the average of their parental breeding values and do not account for Mendelian sampling. They can be thus considered as static. While increasing number of records may improve accuracy of the **A** estimation of parental BVs, it will not exceed a value of 0.7, which is the maximum accuracy given that the BVs of parents are known without error. In contrast, accuracy of genomic predictions is dynamic. Increasing number of genotyped individuals with phenotypic records will increase the accuracy with which QTL effects are estimated, thus reducing the error of predictions (Meuwissen et al., 2001). As the size of the genotyped population grows, the improvement in accuracy will be observed for all individuals. It is therefore expected that with increased sample size, particularly the TRN population, the accuracy of genomic methods will continue to increase.

5.2 Mixing of Matrices

While the best results attained using un-mixed matrices depended on the chip used, mixing of the sources of information proved to bring benefits to both likelihood of the models (for both GWAM and ESM) and accuracy of predictions (particularly visible for ESM chip; lack of this observation on GWAM chip was caused by poor performance of the $\mathbf{G_{LD}}$ matrix). The idea of mixing LD and LA information was first introduced for QTL mapping and revolved around coalescence theory (Meuwissen and Goddard, 2001). Regressing the genomic relationships back to some reference was suggested as a method which would account for

imperfect linkage between markers and QTLs (de los Campos et al., 2013). The method used here resembles the methods used by VanRaden et al (2008) and Goddard et al. (2011), who suggested regressing genomic relationships back to the values expected from the pedigree.

In the presented study, the optimum coefficients varied between λ=0.3 and λ=0.7, depending on the choice of markers. Previous studies in this area presented theoretical calculation of the regression coefficient based on simulated data (Meuwissen and Goddard, 2001, Goddard et al., 2011) however VanRaden et al. (2008) used also an empirical assessment of 4 lambda values. These studies showed that the regression coefficient used for mixing of the matrices depends on the number of markers used and on how well the marker at one locus predicts relationships at other loci (*c*). In fact, *c* has been used as a standalone regression coefficient in a simulation study by de los Campos et al. (2013), where it was calculated as a regression coefficient of marker relationships on the relationships at causal loci. Their results showed that the value of the regression coefficient depended largely on the types of relationships present in the data, with the coefficient calculated for nominally unrelated individuals being considerably lower (between 0.09 - 0.13 depending on the choice of markers), than that calculated for a dataset containing relatives (between 0.28 and 0.37). In practice, estimation of the relationships at causal loci is largely impossible. VanRaden et al. (2008) proposed an approximation to *c*, by taking the error variance of prediction of true fraction of DNA shared between full sibs. Thus, his estimate of the regression coefficient (λ_{VR}) can be presented as:

$$\lambda_{VR} = \frac{0.05^2}{0.05^2 + {}^{0.125}/_{m}}$$

where $(0.05)^2$ is the error variance of prediction of true fraction of DNA shared by two non-inbred fullsibs and ${}^{0.125}/_{m}$ is the error variance of $\mathbf{G_{LD}}$, where *m* is the number of markers used.

Goddard et al. (2011) suggested a similar equation for calculation of the regression coefficient, which can be transformed into an equivalent of the λ_{VR} equation:

$$\lambda_G = 1 - \frac{1}{(cm+1)} = \frac{c}{c + {}^{1}/_{m}}$$

Thus, these two equations, although based on slightly different theoretical assumptions, are very similar and suggest large influence of marker density on the optimum regression coefficient. While both methods of assessing the lambda value tend to 1 as the number of markers increases, the values obtained in other studies based on simulated data varied

between 0.28 (Meuwissen et al., 2011) and 0.78 (Goddard et al., 2011), however, a general tolerance to lambda values was reported as well (Meuwissen et al., 2011).

Substituting *m* for 27,000 markers used in this study to the λ_{VR} equation yields a value of nearly 1, which, according to the theoretical assumptions, is an approximation that should hold for both chips. Using an empirical approach of testing several chosen values we were able to prove that the highest likelihood varies between the types of data, even if marker numbers are the same. Also, the value approximated by the above equation does not agree with lambda value identified with highest likelihood in this data set, and does not fall in the confidence intervals calculated on both chips.

Using the above equation, variable *c* can be estimated by:

$$c = \frac{\lambda/m}{(1-\lambda)}$$

Thus, it cannot be calculated using λ_{VR} estimated for this dataset. Substituting values obtained in this study gives the estimates between 8.64×10^{-5} and 3.17×10^{-5}, for λ=0.7 and λ=0.3 respectively for GWAM and ESM chips, which is considerably lower than the estimate of 0.0035 used by Goddard et al. (2011).

The discrepancy between estimated *c* values is not surprising, as this value is not global for genome and/or sample. Variable *c* depends on the number of independent chromosomal segments segregating in this population (M_e) and thus will vary between subsets of data. The equation given by Goddard et al. (2011) for calculation of the regression coefficient can be reworked to estimate the M_e based on the best estimate of λ. Thus:

$$\lambda = \frac{m}{m+M_e} \qquad \text{becomes} \qquad M_e = \frac{m(1-\lambda)}{\lambda}$$

Substituting lambda values identified in this study as the best fitting data, i.e. in the confidence interval between 0.2 and 0.5 for ESM chip and 0.5 and 0.8 for GWAM chip, gave very high estimates of the number of independent chromosome segments, at 27,000 - 108,000 for ESM chip and 6,750 - 27,000 for GWAM chip. The lower the lambda value used, the higher the estimate of the M_e. These high estimates exceed the estimates obtained using more typical method of estimating the number of independent chromosomal segments. For example, it can be calculated from the equation for accuracy (*r*) given number of records in the training population (*N*) and heritability of the trait (h^2) (Daetwyler et al., 2010)

$$r = \sqrt{\frac{Nh^2}{Nh^2 + M_e}}$$

Substituting values of 0.3 for *r*, 3,146 for *N* and 0.35 for h^2 yields an estimate of M_e=11,133.

An alternative calculation can be obtained by using genome characteristics (Goddard, 2009) as:

$$M_e = \frac{2N_eL}{\log(4N_eL)}$$

where *Ne* is the effective population size and *L* is the length of the genome in Morgans. The estimate of effective population size in broiler chickens is not firmly established, between 50 and 200 (Andreescu et al., 2007). Size of the chicken genome in Morgans is also difficult to estimate, as the recombination rates vary greatly between chromosomes, between 2.1cM/Mb for macrochromosomes and 21cM/Mb for highly recombinant microchromosomes (Hillier et al., 2004). Considering the total size of chicken genome estimated at just over 1Gb (Hillier et al., 2004), these recombination rates lead to estimates of the length of the genome in Morgan varying from 25 to an extreme value of 210 Morgans. Substituting these values gives highly variable estimates of M_e between 293 (N_e=50 and L=25) and 6,981 (N_e=200 and L=210). These values are likely to represent particular parts of the genome. However, even the largest estimate of nearly 7,000 is still far from the estimates calculated using the equation proposed by Goddard et al. (2011) with values of lambda taken from empirical evaluation based on real life data.

Considering those discrepancies, it appears that while utilizing lambda estimated empirically on real data to regress $\mathbf{G_{LD}}$ to $\mathbf{G_{LA}}$ or **A** brings benefits in terms of fit of the models to data and increased accuracy, it is not a good predictor of genomic parameters. The method presented in this study is simplified compared to previously suggested techniques, however, it can be argued that when dealing with real data, choosing the coefficient that fits data best is more suitable than using a single value based on many assumptions, which are not likely to hold for all individuals in the sample.

Compared to the previous methods which concentrated on $\mathbf{G_{LD}}$ and **A** matrices, our study is the first to present the effects of regressing the LD based relationships to LA. The latter is expected to combine benefits of both pedigree and marker information, and thus could be a desirable alternative for **A** as the base of the regression. The structure of the population studied resulted in only minor differences between the two pedigree based matrices

(correlation of 0.98, results not shown), and yet information sourced through $\mathbf{G_{LA}}$ gave higher likelihood and better accuracy than **A** on both chips.

5.3 CHIP COMPARISON

Results obtained using the two chips show a large effect of marker choice, rather than density, on the likelihood and accuracy of predictions, with GWAM chip showing increased likelihood and ESM giving more uniform predictive ability between matrices used. Increased likelihood of $\mathbf{G_{LD}}$ matrix observed in the GWAM chip was expected, as GWAS analysis on the basis of which SNPs were selected for this chip was carried out on the same training population as the variance component estimation. In addition, the GWAM chip was characterized by a larger proportion of low MAF markers than the ESM chip. It has been previously speculated that low MAF markers may explain up to 75% of the genetic variance for BWT (Abdollahi-Arpanahi et al., 2014). Unfortunately, the beneficial fit of $\mathbf{G_{LD}}$ does not seem to be translated well into predictions of breeding values using this in-silico chip, with the regression coefficients indicating considerable over-prediction of GEBVs, with lowest accuracy of all analyses presented recorded for $\mathbf{G_{LD}}$ based on the GWAM chip. This contrast is however not unusual, with the performance of genomic methodology in the inference of variance components frequently departing from its predictive ability (de los Campos, 2014).

As a result of lowered MAF among GWAM chip markers, the $\mathbf{G_{LD}}$ matrix calculated on their basis carried elevated relationship coefficients, as marker information used for calculation of the elements of this matrix is first scaled by $\sqrt{2p(1-p)}$, where p is the frequency of the reference allele. Thus, the elements of the GWAM $\mathbf{G_{LD}}$ matrix were further from the elements of pedigree based matrices (both **A** and **LA**) than elements of the $\mathbf{G_{LD}}$ matrix calculated from the ESM chip, with correlation between off-diagonal elements of $\mathbf{G_{LD}}$ and **A** estimated at 0.3 and 0.61 respectively for GWAM and ESM chips (results not shown). Considering lowered accuracy of predictions based on this matrix and the expectation that as the number of markers increases, the estimates of the genomic relationship matrix should approximate that of **A** matrix, it can be deduced that the $\mathbf{G_{LD}}$ matrix calculated from GWAM markers produced frequency-inflated covariances between individuals.

On the other hand, the effect of SNP choice was less pronounced in LA analyses. As the LA approach concentrates on the information sourced from the last few generations, it can source information from larger chromosome segments. Thus, it can achieve good results even if the marker map is sparse (i.e. even if part of the SNPs in the GWAM chip were not informative for the LD approach, they provided enough background information for LA).

This lower sensitivity of the LA approach was previously confirmed in LA analyses where majority of the information was obtained using linkage analysis when available marker maps were sparse (Meuwissen and Goddard, 2004).

Based on the results of analyses performed on these two chips it can be postulated that while the LA approach is a robust method in respect to the choice of SNPs, wherever possible, evenly spaced markers should be selected to avoid localized overprediction of covariances between relatives in an LD approach. This finding, based on real data, provides a substantial support to results obtained from simulated data, where even spacing of markers provided the most beneficial effect on the accuracy of genomic prediction (Calus et al., 2008)

6. CONCLUSIONS

Analysing real data from a commercial broiler population showed that genomic prediction using relatively low marker densities can improve on the likelihood of models and accuracy of the predictions using pedigree only. The best results are achieved when the relationship matrix used combines different sources of information, with the $\mathbf{G_{LD}}$ matrix regressed back to matrices related to the pedigree structure, i.e. **A** or $\mathbf{G_{LA}}$**.** The optimum regression coefficient in form of lambda has been estimated at λ between 0.3 - 0.7, depending on the choice of markers. These values differ from previously theoretically derived values, which for an equivalent number of markers were speculated to be close to 1. The empirical values of λ found in this study were used to approximate some of the genomic parameters like M_e, however, the results indicate that these empirical values are not suitable for such predictions.

Comparing the results from two different chips showed large influence of the marker choice. Choosing markers based on their effect of the trait in question through GWAS procedures brought a small improvement into the LA approach, however utilizing these markers in the LD approach resulted in biased GEBVs of low accuracy. The chip based on ESM showed more reliable results, with performance of methods following expected patterns.

7. REFERENCES

ABDOLLAHI-ARPANAHI, R., PAKDEL, A., NEJATI-JAVAREMI, A., MORADI SHAHRBABAK, M., MOROTA, G., VALENTE, B. D., KRANIS, A., ROSA, G. J. M. & GIANOLA, D. 2014. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. *Journal of Animal Breeding and Genetics,* 131**,** 183-193.

ANDREESCU, C., AVENDANO, S., BROWN, S. R., HASSEN, A., LAMONT, S. J. & DEKKERS, J. C. 2007. Linkage disequilibrium in related breeding lines of chickens. *Genetics,* 177**,** 2161-9.

CALUS, M. P. L., MEUWISSEN, T. H. E., DE ROOS, A. P. W. & VEERKAMP, R. F. 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics,* 178**,** 553-561.

CARLBORG, Ö., KERJE, S., SCHÜTZ, K., JACOBSSON, L., JENSEN, P. & ANDERSSON, L. 2003. A Global Search Reveals Epistatic Interaction Between QTL for Early Growth in the Chicken. *Genome Research,* 13**,** 413-421.

DAETWYLER, H. D., PONG-WONG, R., VILLANUEVA, B. & WOOLLIAMS, J. A. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics,* 185**,** 1021-31.

DE LOS CAMPOS, G., SORENSEN, D., GIANOLA, D. Genomic Heritability: What is it? 10th World Congress on Genetics Applied to Livestock Production, 18 Aug 2014 2014 Vancouver.

DE LOS CAMPOS, G., VAZQUEZ, A. I., FERNANDO, R., KLIMENTIDIS, Y. C. & SORENSEN, D. 2013. Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genet,* 9**,** e1003608.

DE VERDAL, H., NARCY, A., BASTIANELLI, D., CHAPUIS, H., MEME, N., URVOIX, S., LE BIHAN-DUVAL, E. & MIGNON-GRASTEAU, S. 2011. Improving the efficiency of feed utilization in poultry by selection. 2. Genetic parameters of excretion traits and correlations with anatomy of the gastro-intestinal tract and digestive efficiency. *BMC Genetics,* 12**,** 71.

FALCONER, D. S. & MACKAY, T. F. C. 1996. *Introduction to quantitative genetics / D.S. Falconer and Trudy F.C. Mackay,* Harlow : Longman, 1996.4th ed.

GAYA, L., FERRAZ, J., REZENDE, F., MOURAO, G., MATTOS, E., ELER, J. & MICHELAN FILHO, T. 2006. Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. *Poult Sci,* 85**,** 837-843.

GILMOUR, A. R., GOGEL, B. J., CULLIS, B. R., WELHAM, S. J. & THOMPSON, R. 2006. *{ASREML} User Guide Release 2.0*.

GODDARD, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica,* 136**,** 245-257.

GODDARD, M. E. & HAYES, B. 2007. Genomic Selection. *Journal of Animal Breeding and Genetics,* 124**,** 323 - 330.

GODDARD, M. E., HAYES, B. J. & MEUWISSEN, T. H. E. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of Animal Breeding and Genetics,* 128**,** 409-421.

GRAY, A., STEWART, I. & TENESA, A. 2012. Advanced complex trait analysis. *Bioinformatics,* 28**,** 3134-36.

HABIER, D., FERNANDO, R. L. & DEKKERS, J. C. M. 2007. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics,* 177**,** 2389-2397.

HARTL, D. L. & CLARK, A. G. 1997. *Principles of population genetics,* Sunderland, MA, Sinauer Associates.

HAYES, B. J., VISSCHER, P. M. & GODDARD, M. E. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research,* 91**,** 47-60.

HENDERSON, C. R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,* 31**,** 423-447.

HILLIER, L. W., MILLER, W., BIRNEY, E., WARREN, W., HARDISON, R. C., PONTING, C. P., BORK, P., BURT, D. W., GROENEN, M. A. M., DELANY, M. E., DODGSON, J. B., CHINWALLA, A. T., CLIFTEN, P. F., CLIFTON, S. W.,

DELEHAUNTY, K. D., FRONICK, C., FULTON, R. S., GRAVES, T. A., KREMITZKI, C., LAYMAN, D., MAGRINI, V., MCPHERSON, J. D., MINER, T. L., MINX, P., NASH, W. E., NHAN, M. N., NELSON, J. O., ODDY, L. G., POHL, C. S., RANDALL-MAHER, J., SMITH, S. M., WALLIS, J. W., YANG, S. P., ROMANOV, M. N., RONDELLI, C. M., PATON, B., SMITH, J., MORRICE, D., DANIELS, L., TEMPEST, H. G., ROBERTSON, L., MASABANDA, J. S., GRIFFIN, D. K., VIGNAL, A., FILLON, V., JACOBBSON, L., KERJE, S., ANDERSSON, L., CROOIJMANS, R. P. M., AERTS, J., VAN DER POEL, J. J., ELLEGREN, H., CALDWELL, R. B., HUBBARD, S. J., GRAFHAM, D. V., KIERZEK, A. M., MCLAREN, S. R., OVERTON, I. M., ARAKAWA, H., BEATTIE, K. J., BEZZUBOV, Y., BOARDMAN, P. E., BONFIELD, J. K., CRONING, M. D. R., DAVIES, R. M., FRANCIS, M. D., HUMPHRAY, S. J., SCOTT, C. E., TAYLOR, R. G., TICKLE, C., BROWN, W. R. A., ROGERS, J., BUERSTEDDE, J. M., WILSON, S. A., STUBBS, L., OVCHARENKO, I., GORDON, L., LUCAS, S., MILLER, M. M., INOKO, H., SHIINA, T., KAUFMAN, J., SALOMONSEN, J., SKJOEDT, K., WONG, G. K. S., WANG, J., LIU, B., WANG, J., YU, J., YANG, H. M., NEFEDOV, M., KORIABINE, M., DEJONG, P. J., GOODSTADT, L., WEBBER, C., DICKENS, N. J., LETUNIC, I., SUYAMA, M., TORRENTS, D., VON MERING, C., et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature,* 432**,** 695-716.

KRANIS, A., GHEYAS, A., BOSCHIERO, C., TURNER, F., YU, L., SMITH, S., TALBOT, R., PIRANI, A., BREW, F., KAISER, P., HOCKING, P., FIFE, M., SALMON, N., FULTON, J., STROM, T., HABERER, G., WEIGEND, S., PREISINGER, R., GHOLAMI, M., QANBARI, S., SIMIANER, H., WATSON, K., WOOLLIAMS, J. & BURT, D. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics,* 14**,** 59.

LUAN, T., WOOLLIAMS, J., ODEGARD, J., DOLEZAL, M., ROMAN-PONCE, S., BAGNATO, A. & MEUWISSEN, T. 2012. The importance of identity-by-state information for the accuracy of genomic selection. *Genetics Selection Evolution,* 44**,** 28.

MEUWISSEN, T. & GODDARD, M. 2001. Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution,* 33**,** 605 - 634.

MEUWISSEN, T. & GODDARD, M. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics Selection Evolution,* 36**,** 261 - 279.

MEUWISSEN, T. & GODDARD, M. 2010. The Use of Family Relationships and Linkage Disequilibrium to Impute Phase and Missing Genotypes in Up to Whole-Genome Sequence Density Genotypic Data. *Genetics,* 185**,** 1441-1449.

MEUWISSEN, T., HAYES, B. & GODDARD, M. E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics,* 157**,** 1819 - 1829.

MEUWISSEN, T. H. E., LUAN, T. & WOOLLIAMS, J. A. 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. *Journal of Animal Breeding and Genetics,* 128**,** 429-439.

NEJATI-JAVAREMI, A., SMITH, C. & GIBSON, J. P. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science,* 75**,** 1738-45.

PAYNE, R. W., MURRAY, D. A., HARDING, S. A., BAIRD, D. B. & SOUTAR, D. M. 2009. GenStat for Windows (12th Edition) Introduction. http://www.vsni.co.uk/downloads/genstat/release12/doc/IntroGuide.pdf: VSN International.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. W., DALY, M. J. &

SHAM, P. C. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics,* 81**,** 559-575.

VANRADEN, P. M. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science,* 91**,** 4414-4423.

VILLANUEVA, B., PONG-WONG, R., FERNÁNDEZ, J. & TORO, M. A. 2005. Benefits from marker-assisted selection under an additive polygenic genetic model. *Journal of Animal Science,* 83**,** 1747-1752.

CHAPTER 6

THE EFFECT OF THE TRAINING POPULATION SIZE AND CHIP DENSITY ON ACCURACY AND BIAS OF GENOMIC PREDICTIONS IN BROILER CHICKENS

1. INTRODUCTION

Genomic prediction (GP) is a tool used by livestock breeders to calculate genome-wide estimated breeding values (GEBVs) of selection candidates, based on their marker genotypes. It has been successfully utilized in species such as cattle (Pryce and Daetwyler, 2012) and sheep (Daetwyler et al., 2010) and recently has been gaining recognition in poultry breeding (Avendano et al., 2010). In situations where selection candidates are young animals without phenotypic records, GP can be applied as a two-step procedure, where in the first stage marker effects are calculated using genotypes and phenotypes of a training population (TRN); in the second stage the GEBVs of selection candidates in test population (TST) are calculated based on their own genotypes and marker effects calculated in step 1 (Goddard and Hayes, 2007).

The accuracy with which GEBVs approximate the true breeding values (TBVs) has been found to be dependent on the heritability of the trait, number of genotyped animals with phenotypic records (TRN individuals) and number of independent chromosomal segments segregating in the population (Daetwyler et al., 2008). It depends also on the linkage between markers and quantitative trait loci (QTLs) affecting the trait in question (Goddard and Hayes, 2007). For most traits, the number of QTLs determining trait variance is large, with each locus having a small effect (Hayes and Goddard, 2001). Because the basic framework of GP has been built upon assumption that each QTL is in linkage disequilibrium with at least one marker (Meuwissen et al., 2001), it follows that dense marker maps are needed. Desirable marker densities for accurate genomic predictions of chicken GEBVs reported in the literature vary between 3k (González-Recio et al., 2009) and over 100k SNPs (Megens et al., 2009). The actual marker number required depends on the genetic architecture of the trait and genetic structure of the population.

Among the parameters identified as changing the attainable accuracy, two can be modified through experimental design: marker density and TRN population size. Their effect on

accuracy has to be carefully weighed against the costs of genotyping, particularly in poultry, where the monetary value of an individual selection candidate is low (Avendano et al., 2010). The benefits of increasing the information content through marker density and TRN size have previously been estimated theoretically on simulated data (e.g. Meuwissen, 2009, Goddard, 2009, Daetwyler et al., 2008). These papers have provided a theoretical framework to calculate the desired population size and marker density for accurate estimation of Breeding Values. For example, the size of the TRN population for prediction of GEBVs to achieve high accuracies in excess of 0.7 for unrelated individuals has been estimated as $2N_eL$, where N_e is the effective population size and L is the size of genome in Morgans (Meuwissen, 2009).

The theoretical expectations of minimum marker density and TRN size requirements for accurate GP have been compared to empirical accuracies in species like cattle (e.g. Luan *et al.*, 2009). In cattle, these requirements were found to be in a relatively good agreement with empirical results of milk yield prediction, at 0.46 and 0.39 for predicted and observed accuracies respectively (Luan et al., 2009). Until recently the numbers of chickens genotyped at high density were too low to allow such analyses on real data. Most of the published studies on GS in chickens were limited to fewer than 5k genotyped individuals per line, and genotyped at densities up to around 50k SNPs (Wolc et al. 2010; Chen et al. 2011; Wolc et al. 2011).

The study presented is a first analysis of a large broiler population, consisting of over 23,500 chickens genotyped with high density of over 600k SNP markers. The objective of this study was to establish the effect of the TRN size and chip density on accuracy and bias of genomic predictions of breeding values in broiler population.

2. MATERIALS AND METHODS

2.1 PEDIGREE

The dataset used in the analyses and provided by Aviagen Ltd. was limited to genotyped birds only and contained 23,583 individuals. As the proportion of genotyped selection candidates differed between generations, the depth of the pedigree for this dataset varied between individuals, from 0 for the base population, to a maximum of 8 generations. The pedigree included 441 sires (offspring to 168 paternal grandsires and 231 paternal granddams) and 2351 dams (offspring to 260 maternal grandsires and 640 maternal

granddams). The average number of offspring per sire was 52 (with a maximum of 297) while the average number of offspring per dam was 10 (maximum 52). The base population of the pedigree, i.e. individuals with both sire and dam unknown/non-genotyped consisted of 225 individuals, with a further 703 individuals with one parent unknown (421 with unknown sire and 282 with unknown dam). Due to the experimental nature of the first genotyping experiments at Aviagen Ltd., the base population was not limited to the earliest generations of the dataset. Instead, the individuals with unknown parents were incorporated into the overall population at different points in time.

2.2 PHENOTYPES

The phenotypic measurements of the genotyped birds were obtained for following traits:

- BWT – body weight at 35 days, measured on both males and females, available for all individuals.

- LFI – feed intake between 14 and 35 day, measured in feeding stations on females only, before juvenile selection.

- HHP – the cumulative egg production during the whole laying period, recorded for each hen.

Juvenile body weight was recorded on all birds, preceding the juvenile selection. Female feed intake was recorded prior to juvenile selection as well, however, due to the pre-set definition of this trait, only female individuals in the dataset have been recorded for this trait (n_{LFI}=13,728). Hen housed production, aside from being a sex-limited trait, was recorded mainly on animals that passed three stages of selection, although an outgroup of selection rejects was also included (n_{HHP}=5,419).

Several other traits were also analysed (male feed intake, weight gain, early and late hatchability, early and late fertility) but due to low numbers of records the results were not included in this Chapter, instead they can be found in Appendix 1.

2.3 TRAINING POPULATION

Training population (TRN) is a sample of a studied population, which contains animals with both phenotypes and genotypes, used to calculate marker effects which are later used to estimate BVs of selection candidates (test population - TST), i.e. individuals without their own phenotypic records. The effect of the size of TRN on the accuracy of predictions was assessed through 4 scenarios with different proportions of the population placed in TRN and

TST. First, a pool of potential selection candidates was identified in the data by extracting individuals with no progeny records in the dataset (i.e. descendants and siblings of TRN individuals only). From this group, TST populations of different sizes were created by progressively masking phenotypes of individuals without offspring in the data in such a way, that TST accounted for 40% of the total population in scenario SI, 30% in SII, 20% in SIII and 10% in SIV, with the remaining individuals placed in TRN (contributing 60%, 70%, 80% and 90% of the population respectively).

Assignment of the selection candidates to the TST populations was random in respect to the number of generations behind the selection candidates. Also, the assignment was based solely on the number of animals in the dataset therefore while for BWT the percentages hold for both number of individuals and number of records, for LFI and HHP the numbers of phenotype records do not necessarily agree with the assumed percentage split.

Table 6.1 gives the numbers of phenotypic records available for each trait in each of the scenarios.

Table 6.1 Number of individuals with records for juvenile body weight (BWT), female feed intake (LFI) and hen housed production (HHP) in particular scenarios, presented in the real numbers and as a proportion of all the records for this trait.

Scenario	Data	BWT		LFI		HHP	
		Number	Proportion	Number	Proportion	Number	Proportion
SI	TRN	14,150	60%	7,984	58%	4,167	76%
	TST	9,433	40%	5,744	42%	1,252	24%
SII	TRN	16,508	70%	9,632	70%	4,535	84%
	TST	7,075	30%	4,096	30%	884	16%
SIII	TRN	18,866	80%	10,820	78%	4,828	89%
	TST	4,717	20%	2,908	22%	591	11%
SIV	TRN	21,225	90%	12,276	89%	5,150	95%
	TST	2,358	10%	1,452	11%	269	5%
Total		23,583		13,729		5,420	

2.4 GENOTYPES

Genotypes of the birds used in the analyses were obtained using the full range of SNP markers on Affymetrix Axiom 600K panel, spread over 28 autosomes, and sex chromosome Z (Kranis et al., 2013). Out of the 23,583 birds, 1,446 were genotyped using this high density panel, while the 600K genotypes for the rest were imputed from a low density panel containing 3k SNPs. The imputation was performed by A. Kranis (Aviagen Ltd.) using AlphaImpute (Hickey et al., 2011), and based on a validation sample analysis, was characterised by high accuracy of 0.97.

2.5 CHIP DENSITY

To evaluate the effect of chip density on accuracy of prediction, 6 additional chips were extracted from the total number of SNPs available. Starting from the smallest chip, each consecutive chip was created by adding a random sample of SNPs to the markers contained in the previous, smaller chip.

After the SNP selection, each chip was subjected to quality control (QC) carried out in Plink software (Purcell et al., 2007). The QC involved removing markers that were not mapped to known chromosomes, markers placed on sex chromosomes and markers which did not pass the threshold values for pre-set screening criteria. The screening criteria used were Hardy Weinberg equilibrium (exclude markers which failed Hardy Weinberg Equilibrium test with p<=0.001), missing markers (exclude markers with more than 5% genotypes missing) and minor allele frequency (exclude markers with MAF<0.01). Out of the 28 autosomes with markers, chromosome 16 was represented by a lower number of markers and none of them passed the QC thus resulting in the final dataset represented by markers spread over 27 autosomes. The number of SNPs before and after the quality control, listed against the selection criteria is given in Table 6.2.

Table 6.2 Quality control statistics of the 7 chips used in the analyses, with the number of markers that failed QC expressed as a percentage of the total number of mapped markers. Column Z gives the number of markers on sex chromosome Z.

Chip name	Mapped markers	Markers failed due to:			Remaining
		HWE p<=0.001	Missingness > 0.05	MAF<0.01	
2k	3 072	7%	7%	11%	2 337
7k	9 992	8%	6%	11%	7 606
19k	24 978	7%	6%	11%	19 019
40k	52 618	8%	6%	11%	40 052
70k	99 542	6%	6%	18%	70 612
134k	198 730	5%	6%	22%	134 924
412k	625 995	5%	6%	25%	412 692

2.6 BREEDING VALUE PREDICTION

Breeding values of TST individuals were predicted using the phenotypes and genotypes of their relatives in TRN population and the relationships they shared. This was achieved by fitting mixed linear models in ACTA software package (Gray et al. 2012) for genomic predictions (GBLUP) and in ASReml for pedigree based (PED) predictions (Gilmour et al., 2006). Using two separate software packages was necessary due to the computational

limitation of both ACTA and ASReml, which was discussed earlier. ACTA software was unable to predict the pedigree-based EBVs for individuals without phenotypes, while ASReml was unable to cope with genomic predictions on such large amounts of data. However, the two packages have produced similar estimates of variance components on the same small dataset, which gives confidence in comparing the results.

The statistical assumptions behind the MLM were defined previously. The **G** matrix used for the analyses was constructed in the ACTA software package (Gray et al., 2012) which follows methodology defined by Yang et al. (2011) and which has been outlined in Chapter 3. Due to different markers used, the **G** matrices varied between chip densities.

The random effect fitted for all traits was the effect of the animal. The fixed effect fitted to BWT and LFI was coded as 'hwumgs', spread over 2390 levels with average 13 chicks per level (between 1 and 151), which combined several environmental and husbandry factors (hatch week, unit, mating group and sex). The fixed effect fitted to HHP was limited to hatch week. The latter effect had 225 levels, with average 104 chickens in each (min 1, max 679).

Each of the mixed linear model analyses was iterated until convergence, i.e. until variance estimates changed by less than 1% between iterations and likelihood estimates changed by less than 0.002 x *I*, where *I* is the current convergence number (Gilmour et al., 2006).

Accuracies of phenotype (r_P) and breeding value (r_A) predictions were calculated as outlined in Chapter 4. The bias of prediction was calculated in GenStat as the regression coefficient of phenotypes regressed on the predicted (G)EBVs, with fixed effects accounted for in the model.

3. RESULTS

3.1 VARIANCE COMPONENTS

The estimates of the variance components of the three traits are presented in Tables 6.3, 6.4 and 6.5 for BWT, LFI and HHP respectively.

For PED, the estimate of error variance of BWT was similar for SII, SIII and SIV, with a slight decrease in SI. The estimates of genetic variance were also similar across the scenarios. This resulted in heritability estimates being nearly equal for SII, SIII and SIV at 0.41, with peak estimated from SI at 0.47. This peak was caused more by lower error variance estimate, than higher estimate of genetic variance in SI.

A considerable proportion (between 27 and 48%) of the genetic variance identified by PB was not detected by genomic methods, which resulted in increased estimates of error variance and decreased estimates of genetic variance. This was observed for all chips, particularly for low density chips. Increasing the density of markers from 2k to 70k brought about a decrease in the estimates of the error variance, while increasing the density further did not seem to bring additional reduction. Similarly, the estimates of genetic variance plateaued when density exceeded 70k markers, after initial increase for chips 2k to 70k. The change in the partitioning of the variance with the chip density brought about an increase in heritability, however the differences were minor for all chips except the 2k chip.

Table 6.3 Estimates of BWT variance components of pedigree based (PB) and genomic analyses run on different chips and with different splits of data into TRN and TST. σ^2_A - variance of the direct genetic effect of an individual, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ^2_A (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)
PB	SI	61.15 (4.76)	87.55 (2.85)	148.70 (2.82)	0.47 (0.03)
	SII	60.45 (4.42)	90.22 (2.66)	150.67 (2.61)	0.40 (0.02)
	SIII	62.59 (4.25)	90.53 (2.54)	153.12 (2.50)	0.41 (0.02)
	SIV	62.69 (4.00)	90.66 (2.39)	153.35 (2.36)	0.41 (0.02)
2k	SI	34.39 (2.34)	109.18 (1.49)	143.57 (2.57)	0.30 (0.01)
	SII	32.27 (2.12)	112.46 (1.4)	144.73 (2.37)	0.22 (0.01)
	SIII	33.05 (2.07)	113.87 (1.31)	146.91 (2.29)	0.22 (0.01)
	SIV	32.40 (1.96)	114.69 (1.23)	147.09 (2.18)	0.22 (0.01)
7k	SI	39.01 (2.49)	104.93 (1.5)	143.94 (2.57)	0.27 (0.01)
	SII	37.72 (2.3)	107.90 (1.4)	145.62 (2.39)	0.26 (0.01)
	SIII	38.79 (2.22)	109.20 (1.31)	147.99 (2.32)	0.26 (0.01)
	SIV	38.49 (2.11)	109.75 (1.22)	148.24 (2.21)	0.26 (0.01)
19k	SI	41.54 (2.62)	103.37 (1.51)	144.90 (2.62)	0.29 (0.01)
	SII	40.76 (2.44)	106.12 (1.41)	146.88 (2.46)	0.28 (0.01)
	SIII	41.46 (2.34)	107.57 (1.32)	149.03 (2.36)	0.28 (0.01)
	SIV	41.44 (2.24)	108.10 (1.23)	149.53 (2.27)	0.28 (0.01)
40k	SI	43.50 (2.7)	102.24 (1.51)	145.73 (2.66)	0.30 (0.01)
	SII	42.12 (2.5)	105.29 (1.41)	147.41 (2.48)	0.29 (0.01)
	SIII	43.00 (2.4)	106.66 (1.32)	149.66 (2.39)	0.29 (0.01)
	SIV	42.96 (2.3)	107.21 (1.23)	150.17 (2.29)	0.29 (0.01)
70k	SI	44.46 (2.75)	101.93 (1.52)	146.39 (2.68)	0.30 (0.01)
	SII	43.13 (2.54)	104.93 (1.42)	148.06 (2.5)	0.29 (0.01)
	SIII	44.20 (2.45)	106.24 (1.32)	150.44 (2.42)	0.29 (0.01)
	SIV	44.29 (2.35)	106.79 (1.24)	151.08 (2.32)	0.29 (0.01)
134k	SI	45.43 (2.8)	101.64 (1.53)	147.08 (2.71)	0.31 (0.01)
	SII	44.17 (2.59)	104.62 (1.42)	148.78 (2.53)	0.30 (0.01)
	SIII	45.31 (2.5)	105.92 (1.33)	151.23 (2.45)	0.30 (0.01)
	SIV	45.41 (2.39)	106.51 (1.24)	151.92 (2.35)	0.30 (0.01)
412k	SI	45.83 (2.82)	101.56 (1.53)	147.40 (2.72)	0.31 (0.01)
	SII	44.62 (2.61)	104.48 (1.42)	149.10 (2.55)	0.30 (0.01)
	SIII	45.86 (2.52)	105.73 (1.33)	151.59 (2.46)	0.30 (0.01)
	SIV	45.74 (2.41)	106.40 (1.24)	152.15 (2.36)	0.30 (0.01)

For BWT analyses using marker data, increase in size of the TRN caused an increase in the estimates of error variance, with a consistent pattern observed across chip densities. The changes of the genetic variance estimates were less directional, with only minor fluctuations. Thus, an increase in error variance with TRN size brought about a decrease in heritability; however the differences for this parameter were negligible for most chip densities.

For female feed intake, increasing number of records available in PB predictions across splits did not bring a clear pattern of change in the estimates of variance components (Table 6.4). Similarly to BWT, LFI PB variance inference was more efficient at identifying genetic variance than genomic methods, with the latter having a larger proportion of variance identified as residual.

Increasing the size of TRN from SI to other scenarios brought about an increase in LFI error variance estimates. For the other three scenarios the estimates fluctuated around the same value. For the estimates of genetic variance this trend was reversed, with estimates obtained using SI being higher than from other scenarios. Due to the small magnitude of those differences, the changes in heritability across different TRN sizes were negligible.

Increasing chip density in LFI analyses brought about a steady decrease in the estimate of error variance for chips between 2k and 40k, with diminishing increase observed for higher densities. The estimate obtained from the 70k chip nearly reached a plateau and remained stable for chips 134k and 431k. In contrast, the estimate of genetic variance increased steadily across the range of chips, although the differences at high density chips (134k and 412k) were minor. Thus, the heritability estimates using marker data ranged between 0.29 and 0.42 for chips 2k and 412k respectively.

Table 6.4 Estimates of LFI variance components of pedigree based (PB) and genomic analyses run on different chips and with different splits of data into TRN and TST. σ^2_A - variance of the direct genetic effect of an individual, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ^2_A (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)
PB	SI	137.38 (11.72)	146.13 (7.07)	283.5 (6.7)	0.48 (0.03)
	SII	141.39 (12.44)	143.67 (7.44)	285.05 (7.1)	0.5 (0.03)
	SIII	137.66 (11.19)	147.79 (6.74)	285.45 (6.39)	0.48 (0.03)
	SIV	139.7 (10.67)	148.25 (6.4)	287.95 (6.09)	0.49 (0.03)
2k	SI	90.00 (6.57)	188.80 (3.58)	278.80 (6.77)	0.32 (0.02)
	SII	81.40 (5.7)	195.12 (3.28)	276.52 (6.0)	0.29 (0.02)
	SIII	80.62 (5.43)	197.50 (3.09)	278.12 (5.74)	0.29 (0.01)
	SIV	79.03 (5.12)	198.10 (2.87)	277.13 (5.44)	0.29 (0.01)
7k	SI	103.66 (7.09)	177.06 (3.6)	280.73 (6.85)	0.37 (0.02)
	SII	97.47 (6.29)	182.13 (3.27)	279.60 (6.17)	0.35 (0.02)
	SIII	98.00 (6.02)	184.07 (3.07)	282.07 (5.93)	0.35 (0.02)
	SIV	97.22 (5.68)	184.13 (2.83)	281.35 (5.63)	0.35 (0.01)
19k	SI	110.18 (7.41)	172.60 (3.63)	282.78 (6.98)	0.39 (0.02)
	SII	102.45 (6.53)	178.46 (3.29)	280.91 (6.25)	0.36 (0.02)
	SIII	102.87 (6.25)	180.41 (3.1)	283.28 (6.0)	0.36 (0.02)
	SIV	103.03 (5.94)	180.32 (2.86)	283.35 (5.73)	0.36 (0.01)
40k	SI	114.16 (7.59)	170.03 (3.63)	284.19 (7.07)	0.40 (0.02)
	SII	106.46 (6.72)	176.14 (3.3)	282.60 (6.34)	0.38 (0.02)
	SIII	106.51 (6.4)	178.11 (3.1)	284.62 (6.07)	0.37 (0.02)
	SIV	106.03 (6.04)	178.31 (2.86)	284.34 (5.78)	0.37 (0.01)
70k	SI	118.81 (7.83)	168.91 (3.65)	287.72 (7.24)	0.41 (0.02)
	SII	109.87 (6.88)	175.28 (3.32)	285.15 (6.45)	0.39 (0.02)
	SIII	109.93 (6.55)	177.18 (3.11)	287.10 (6.17)	0.38 (0.02)
	SIV	109.27 (6.19)	177.52 (2.87)	286.78 (5.87)	0.38 (0.02)
134k	SI	121.19 (7.98)	168.54 (3.66)	289.73 (7.34)	0.42 (0.02)
	SII	112.47 (7.04)	174.83 (3.33)	287.29 (6.56)	0.39 (0.02)
	SIII	112.94 (6.71)	176.52 (3.12)	289.47 (6.29)	0.39 (0.02)
	SIV	112.54 (6.35)	176.75 (2.88)	289.28 (5.99)	0.39 (0.02)
412k	SI	122.68 (8.06)	168.30 (3.67)	290.98 (7.41)	0.42 (0.02)
	SII	113.80 (7.11)	174.62 (3.34)	288.42 (6.61)	0.39 (0.02)
	SIII	114.18 (6.77)	176.32 (3.13)	290.50 (6.33)	0.39 (0.02)
	SIV	113.36 (6.38)	176.66 (2.89)	290.02 (6.01)	0.39 (0.02)

Similarly to LFI, variance component estimates for HHP estimated through PB did not show a discernible pattern of change with variable TRN size (Table 6.5). The proportion of variance identified by PB as genetic far exceeded the proportion of genetic variance estimated using marker data, particularly at low marker densities, resulting in higher PB estimates of heritabilities, as has been observed for LFI and BWT.

Table 6.5 Estimates of HHP variance components of pedigree based (PB) and genomic analyses run on different chips and with different splits of data into TRN and TST. σ_A^2 - variance of the direct genetic effect of an individual, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ_A^2 (SE)	σ_e^2 (SE)	σ_P^2 (SE)	h^2 (SE)
PB	SI	203.77 (27.62)	510.12 (20.97)	713.89 (18.38)	0.29 (0.03)
	SII	200.92 (27.94)	506.96 (21.24)	707.88 (18.58)	0.28 (0.04)
	SIII	204.25 (27.04)	511.56 (20.5)	715.81 (18.0)	0.29 (0.03)
	SIV	199.48 (26.27)	522.37 (20.06)	721.85 (17.54)	0.28 (0.03)
2k	SI	85.65 (16.13)	620.02 (19.34)	705.66 (20.31)	0.12 (0.02)
	SII	81.31 (14.86)	626.74 (18.35)	708.04 (19.29)	0.11 (0.02)
	SIII	83.20 (14.38)	626.10 (17.57)	709.30 (18.67)	0.12 (0.02)
	SIV	83.85 (13.69)	627.83 (16.83)	711.68 (18.06)	0.12 (0.02)
7k	SI	102.62 (18.32)	604.54 (19.85)	707.16 (20.51)	0.15 (0.02)
	SII	93.30 (16.51)	615.13 (18.83)	708.43 (19.35)	0.13 (0.02)
	SIII	93.88 (15.9)	615.90 (18.04)	709.78 (18.71)	0.13 (0.02)
	SIV	94.80 (15.27)	618.01 (17.31)	712.81 (18.14)	0.13 (0.02)
19k	SI	104.83 (18.85)	602.82 (20.07)	707.64 (20.54)	0.15 (0.15)
	SII	99.76 (17.38)	609.76 (19.0)	709.52 (19.48)	0.14 (0.14)
	SIII	98.81 (16.63)	611.89 (18.23)	710.79 (18.81)	0.14 (0.14)
	SIV	101.21 (16.11)	612.95 (17.49)	714.16 (18.27)	0.14 (0.14)
40k	SI	109.05 (19.38)	599.02 (20.16)	708.06 (20.63)	0.15 (0.03)
	SII	102.89 (17.78)	606.81 (19.09)	709.69 (19.53)	0.14 (0.02)
	SIII	101.95 (17.01)	609.05 (18.3)	711.00 (18.86)	0.14 (0.02)
	SIV	103.87 (16.45)	610.49 (17.56)	714.36 (18.32)	0.15 (0.02)
70k	SI	109.40 (19.69)	600.13 (20.31)	709.53 (20.71)	0.15 (0.03)
	SII	103.17 (18.04)	607.91 (19.21)	711.08 (19.6)	0.15 (0.02)
	SIII	103.16 (17.34)	609.48 (18.42)	712.65 (18.95)	0.14 (0.02)
	SIV	104.67 (16.72)	611.33 (17.67)	716.00 (18.39)	0.15 (0.02)
134k	SI	110.14 (19.98)	600.70 (20.39)	710.84 (20.8)	0.15 (0.03)
	SII	104.28 (18.32)	608.10 (19.28)	712.37 (19.68)	0.15 (0.02)
	SIII	104.33 (17.63)	609.74 (18.49)	714.07 (19.04)	0.15 (0.02)
	SIV	105.96 (17.02)	611.68 (17.75)	717.63 (18.49)	0.15 (0.02)

412k	SI	112.16 (15.51)	577.07 (15.72)	689.23 (16.74)	0.16 (0.02)
	SII	110.43 (14.74)	586.31 (15.14)	696.73 (16.18)	0.16 (0.02)
	SIII	111.07 (14.43)	592.57 (14.75)	703.63 (15.87)	0.16 (0.02)
	SIV	113.33 (14.14)	596.77 (14.3)	710.11 (15.57)	0.16 (0.02)

Increasing the size of TRN in marker-based analyses of HHP caused a slight increase in error variance estimates, particularly for SII, across chip densities. The estimates of genetic variance were highest for SI, with resulting highest estimate of heritability. Compared to BWT and LFI, the standard errors of the HHP estimates were larger, due to considerably lower number of phenotypic records available for this trait.

3.2 ACCURACY OF BV PREDICTION

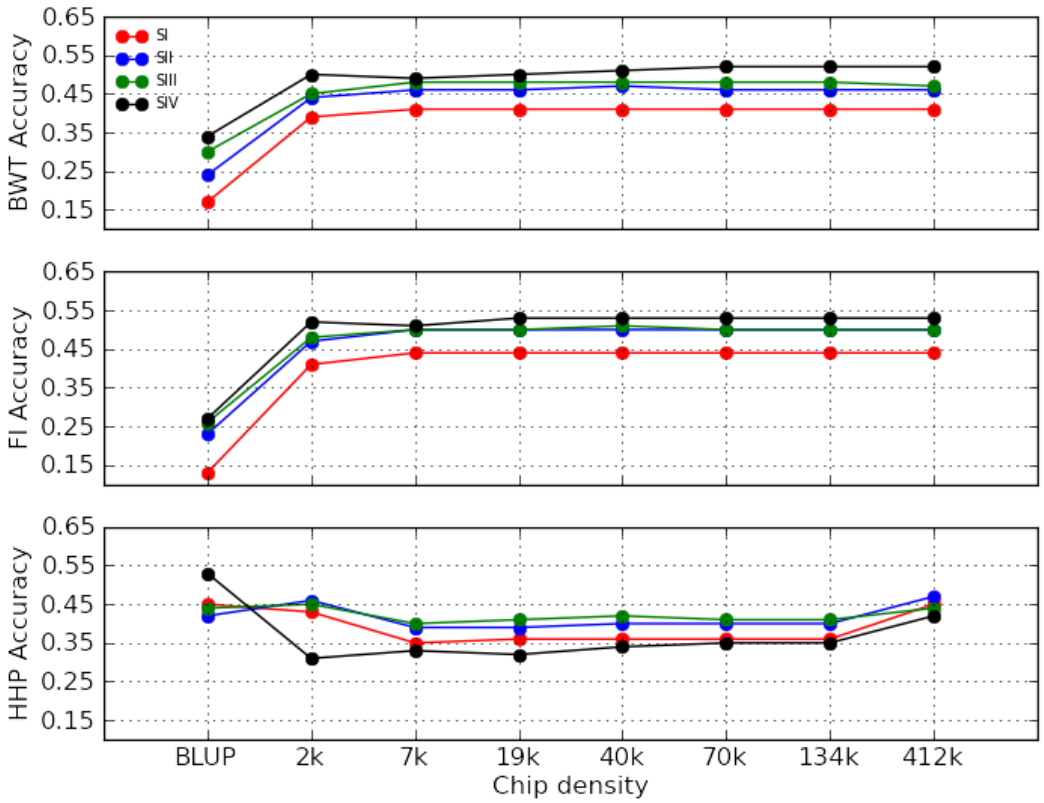

Figure 6.1 Profile of change in accuracy brought about by changes in TRN size and chip density.

The profile of changes in accuracy with variable TRN population size and chip densities for the three traits is presented in Figure 6.1. At the highest chip density and largest TRN size considered here, the magnitudes of the genomic accuracies observed ranged from 0.42 for HHP to 0.53 for LFI. The rankings of accuracy among the traits increased as expected with Nh^2, i.e. HHP < BWT < LFI.

For BWT and LFI the effect of increasing N was found to have a consistent effect in increasing the accuracy of prediction (Figure 6.1). For HHP, a similar trend was observed for SI, SII and SIII, however, the accuracy of SIV dropped below the estimates of SI (Figure 6.1).

In contrast increasing chip density had only a small effect on BWT and LFI accuracy, with negligible increases above 19k density. Yet again, results for HHP were less consistent than for BWT and LFI, and the increased standard errors associated with the HHP accuracies

made comparisons across scenarios difficult. However, in SIV increasing chip density appeared to lead to increased accuracy.

3.3 BIAS OF BV PREDICTION

The estimates of bias for BWT (G)EBV predictions are given in Table 6.6, for LFI in Table 6.7 and for HHP in Table 6.8 For BWT, all predictions overestimated the range of true breeding values, as shown by the regression coefficients being consistently smaller than 1. The lowest bias across different predictions was found for GBLUP using SIV, at β=0.89 (SE 0.06). Increasing the size of TRN was found to reduce the amount of bias for both pedigree and marker based methods, with predictions based on SI having the largest and predictions based on SIV having the lowest bias for both methods. For PED, the difference between regression coefficients obtained in SI and SIV reached significance at t=3.07 ($p<0.01$), while for genomic methods the statistical significance of the difference fluctuated just below the threshold value of $p<0.01$ (t=1.93 for 2k chip and t=1.88 for 412k, both characterized by p=0.03). Although the trend was similar between PED and GBLUP, the latter resulted in predictions with lower bias. Thus, the most biased of the genomic predictions (i.e. GBLUP predictions using SI, resulting in β=0.76) were comparable to the least biased pedigree predictions (i.e. PED predictions using SIV, resulting in β=0.79). The density of chips used in genomic predictions did not have an effect on the bias.

Table 6.6 Estimates of bias in BWT BV prediction using different chip densities and different splits of the population into TRN and TST.

Chip	SI	SII	SIII	SIV
PED	0.50 (0.05)	0.62 (0.05)	0.75 (0.06)	0.79 (0.08)
2k	0.76 (0.03)	0.85 (0.04)	0.83 (0.04)	0.89 (0.06)
7k	0.77 (0.03)	0.84 (0.03)	0.84 (0.04)	0.84 (0.05)
19k	0.77 (0.03)	0.84 (0.03)	0.84 (0.04)	0.85 (0.05)
40k	0.76 (0.03)	0.85 (0.03)	0.84 (0.04)	0.86 (0.05)
70k	0.77 (0.03)	0.85 (0.03)	0.84 (0.04)	0.88 (0.05)
134k	0.76 (0.03)	0.85 (0.03)	0.84 (0.04)	0.88 (0.05)
412k	0.77 (0.03)	0.85 (0.03)	0.83 (0.04)	0.88 (0.05)

As for BWT, predictions of LFI BVs performed using PED were significantly affected by the size of TRN population (t=2.9, $p<0.01$), with the amount of bias almost halved between SI and SIV (Table 6.7). The difference between bias of genomic predictions using SI and SIV was also significant for 2k chip (t=2.5, $p<0.01$), but it failed to reach significance threshold for other chip densities. The bias of genomic predictions of LFI was always lower than that of pedigree based predictions (Table 6.7). Genomic predictions however also

showed sensitivity to the TRN size. The overall bias of LFI predictions using GBLUP was similar to that of BWT predictions, with the best predictions characterised by the same regression coefficient, i.e. β=0.89 (SE 0.06) calculated for GBLUP in SIV.

Table 6.7 Estimates of bias in LFI BV prediction using different chip densities and different splits of the population into TRN and TST.

Chip	SI	SII	SIII	SIV
PED	0.36 (0.05)	0.64 (0.06)	0.65 (0.07)	0.66 (0.09)
2k	0.72 (0.03)	0.85 (0.04)	0.85 (0.05)	0.89 (0.06)
7k	0.74 (0.03)	0.84 (0.04)	0.81 (0.04)	0.81 (0.06)
19k	0.74 (0.03)	0.83 (0.04)	0.82 (0.04)	0.83 (0.06)
40k	0.73 (0.03)	0.84 (0.04)	0.82 (0.04)	0.84 (0.06)
70k	0.73 (0.03)	0.85 (0.04)	0.82 (0.04)	0.85 (0.06)
134k	0.74 (0.03)	0.86 (0.04)	0.83 (0.04)	0.84 (0.06)
412k	0.74 (0.03)	0.86 (0.04)	0.83 (0.06)	0.85 (0.06)

In contrast to BWT and LFI, the range of pedigree based HHP EBVs was underestimated, as shown in Table 6.8. Genomic analyses yielded nearly unbiased estimates for splits SII and SIII, whereas SIV gave predictions with larger bias. Due to the low number of individuals for which the regression was calculated in this trait, the regression coefficient of HHP was accompanied by large standard errors. Although comparisons of the results obtained from different chips is difficult due to the magnitude of the standard errors, it appears that for SIV increasing marker density reduced the amount of bias. For other scenarios the values found for different chips fluctuated without a clear pattern.

Table 6.8 Estimates of bias in HHP BV prediction using different chip densities and different splits of the population into TRN and TST.

Chip	SI	SII	SIII	SIV
PED	1.32 (0.15)	1.10 (0.16)	1.29 (0.22)	1.08 (0.23)
2k	1.07 (0.13)	1.11 (0.15)	1.01 (0.17)	0.55 (0.21)
7k	0.88 (0.13)	0.99 (0.16)	0.94 (0.18)	0.61 (0.21)
19k	0.92 (0.13)	0.98 (0.16)	0.98 (0.18)	0.60 (0.21)
40k	0.89 (0.13)	0.98 (0.15)	0.99 (0.18)	0.64 (0.21)
70k	0.94 (0.14)	1.04 (0.16)	1.02 (0.19)	0.68 (0.22)
134k	0.96 (0.14)	1.05 (0.16)	1.03 (0.19)	0.71 (0.23)
412k	1.07 (0.12)	1.05 (0.14)	0.97 (0.17)	0.72 (0.20)

4. DISCUSSION

The main objective of this study was to evaluate the numbers of TRN individuals and marker densities needed for accurate prediction of GEBVs in broiler populations. Estimation of these two parameters is crucial for a successful implementation of GS in poultry, where the

value of a single selection candidate is low, as compared to the costs of genotyping. The accuracy of genomic predictions found in the analysed dataset showed great promise for the use of GS in broilers, with prediction accuracy of BWT increased by 24% and LFI accuracy increased by 77%, as compared to predictions based on pedigree only. Genomic methods also seemed to cope with selection bias better than BLUP PED. The accuracy of genomic predictions is expected to improve further with growing numbers of genotyped and phenotyped individuals, as the presented results showed the increase in accuracy with increasing TRN size. This effect of TRN population size was found to exceed the effect of increasing marker density, with accuracies of predictions at a given TRN size reaching a plateau at around 20k SNP markers.

Until recently, the numbers of genotyped broilers were too low to evaluate the empirical benefits of increasing TRN in chickens. The size of the TRN populations used in the presented analyses (between 14,150 in SI and 21,225 in SIV for BWT) far exceeds the numbers previously reported, none of which exceeded 3,500 genotyped individuals per line (Andreescu et al., 2010, Chen et al., 2011, Simeone et al., 2012, Wolc et al., 2010, Wolc et al., 2011a) . The beneficial effect of increasing TRN size on accuracy of predictions has been shown in the first study introducing GS (Meuwissen et al., 2001). The authors speculated that the advantages of including more individuals in TRN arise from reduction of the sampling error on the estimates of the marker effects. The presented results provide an empirical confirmation of the expected increase in accuracy with increasing TRN size in broiler chickens. Although the numbers of individuals in TRN were already large even in the SI, particularly for BWT and LFI at 14,150 and 7,984 respectively, increasing them further significantly improved the accuracy of predictions. It is possible that a part of the observed improvement was attributable to a decrease in TST size between the scenarios if there were significant differences between the TST individuals, nevertheless, this has been addressed during the calculation of the accuracy, where both the phenotypes and (G)EBVs were corrected for fixed effects, thus hopefully removing trends and other possible genetic differences among different contemporary groups.

Based on a simulation study, Meuwissen (2009) suggested that for a trait with heritability of 0.8, accuracy of ~0.9 can be achieved when the TRN population contains at least $2N_eL$ individuals, where N_e is the effective population size and L is the size of genome in Morgans (Meuwissen, 2009). The estimate of L depends on the recombination rates, which vary greatly across chicken genome (Hillier et al., 2004). Considering the theory of obligatory cross-overs required for correct segregation of chromosomes during meiosis, it follows that

microchromosomes of chicken genome will be characterised by elevated recombination rates (Rodionov, 1996). Microchromosomes have also been found to have increased GC content, CpG island density, gene density and length and repeat density (Hillier et al., 2004). As a result, the recombination rates across chicken genome have been found to span between 2.5cM/Mb for microchromosomes and 21cM/Mb for macrochromosomes (Hillier et al., 2004). Elevated recombination rates of microchromosomes thus result in their linkage lengths oscillating between 50-100cM (Hillier et al., 2004). Using the quoted recombination rates and the physical length of chicken genome estimated at 1.05Gb (Hillier et al., 2004), results in the length of chicken genome estimated between 25M and 210M. While the latter value is clearly an overestimation, L=25M could be treated as the lower limit of the genome size. The actual size is expected to be somewhat larger. A recent estimate of L based on data from 10K SNP markers spread over 34 linkage groups falls close to it, at L=32M (Groenen et al., 2009). However, the same authors note that their data was missing at least 5 chromosome pairs. Thus, the expected size of the complete chicken genome was estimated at 37M (Groenen et al., 2009). In another study, the length of the genome in chickens was estimated at 30M for sex average map based on a linkage map built using 18K markers spread over 31 linkage groups (Elferink et al., 2010). In the same study, a considerable difference in genome length was observed between different broiler lines, spanning between 28 and 35M. The difference was caused mostly by differences in location of recombination hot spots, and to some degree by informative markers segregating in the two populations (Elferink et al., 2010). This shows that aside from the complex nature of the chicken genome itself, the estimates of the length of the genome are also likely to vary between studied populations. Similarly, the estimates of the effective population size in broiler chickens vary between lines, ranging between 50 and 200 (Andreescu et al., 2007). The N_e for the line used in this study is estimated at 100 (A. Kranis, personal communication).

Taking the minimum (L=25M) and Groenen et al. (2009) estimate of L (L=37M) results in the $2N_eL$ estimated between 5,000 and 7,400, when N_e=100. These numbers were achieved for BWT and LFI across the scenarios, but were just met for HHP (N=5,150 in SIV). However, the heritability of traits studied here was less than 0.8, at $h^2_{BWT} = 0.41$, $h^2_{FI} = 0.48$ and $h^2_{HHP} = 0.28$. Since accuracy is a function of the product of TRN size and h^2 (Daetwyler et al. 2008) this would suggest a TRN size of $1.6N_eL/h^2$ (as $2N_eL = \frac{1.6N_eL}{0.8}$, where 0.8 was the heritability assumed by Meuwissen, 2009); resulting in values of 14,439 for BWT, 12,333 for LFI and 21,143 for HHP, for L= 37M and N_e=100. These numbers were nearly achieved in SI and exceeded in other scenarios for BWT, nearly met for LFI in SIV

and not at all for HHP. Despite the fact that the numbers required were exceeded for BWT, the expected accuracy of 0.9 was still not achieved. Although the numbers required were not met for LFI, it is extremely unlikely that adding the missing 57 individuals (the difference between required 12,333 and 12,276 available in the data) would result in accuracy increasing from observed 0.53 to 0.9. Furthermore, accuracies obtained using the analysed dataset should exceed accuracies predicted using this equation due to the abundance of family relationships present, as the approximation of $2N_eL$ was proposed for populations containing unrelated individuals (Meuwissen, 2009).

Substituting N_e=200, which is an upper limit of currently available estimates of effective population size in broiler lines (Andreescu et al., 2007) in $1.6N_eL/h^2$ (at L=37) yields estimates of 28,878, 24,667 and 42,286 for BWT, LFI and HHP respectively. These values exceed the numbers available in this study for all scenarios and all traits, particularly for HHP.

The accuracies obtained in this study can be also related to the deterministic prediction of accuracy proposed by Daetwyler *et al.* (2008):

$$r_D = \sqrt{Nh^2/(Nh^2 + M_e)}$$

where N is the number of individuals in TRN and M_e is the number of independent segments segregating in the population, with expectation of $M_e = 2N_eL/\log(4N_eL)$, following Goddard *et al* (2009). Substituting N with the SIV TRN size (N_{BWT}=21,225 and N_{LFI}=12,276 N_{HHP}=5,150), h^2 with previously listed heritability estimates yields r_D exceeding 0.9 for BWT and LFI across all M_e values and between 0.71-0.77 for HHP, depending on values substituted for N_e and L (N_e between 50 and 200; L between 25M and 37M). Thus, in theory, the numbers of individuals in TRN available in this study should be large enough for nearly perfectly accurate predictions of GEBVs. This was not found here. It appears that the estimates of M_e obtained using the above parameters of N_e and L are relatively low, with the estimates obtained using the available parameters for broiler populations spanning between M_e=294 (N_e=50 and L=25M) and M_e=1,438 (N_e=200 and L=37M). It has been suggested that although M_e is directly related to the effective population size, it could be referring to historical, rather than the present population size (Hayes et al., 2003). Thus, M_e could be larger than would be predicted from current estimates of N_e. According to de los Campos et al. (2013), due to differences in the M_e estimates obtained using different methods, the use of this term in deterministic predictions of the accuracy is controversial.

Nevertheless, the same equation could be transformed into an approximation of numbers of individuals needed for a given accuracy, as:

$$N = \frac{r^2 M_e}{h^2(1 - r^2)}$$

In scenarios where M_e=771 is accepted (N_e=100 and L=37M), the number of individuals required for r=0.9 would be N_{BWT}=8,013, N_{LFI}=6,845 and N_{HHP}=11,733. These numbers have been far exceeded for BWT and LFI in this study with accuracy observed far from expected 0.9. Similarly, using the larger estimate of M_e=1,438 calls for numbers of individuals (N_{BWT}=14,947, N_{LFI} =12,767 and N_{HHP}=21,887) that were met in this study for BWT and nearly LFI, but failed to achieve the desired r=0.9.

Considering the above, the deterministic predictions of required numbers of individuals in TRN thus found no confirmation in results of this study. There are several possible explanations for this discrepancy. A possible explanation to this reduction is weak LD between markers and QTLs observed in this dataset (Meuwissen et al., 2001). However, given the lack of marked improvement with increasing marker density above 19k suggests that this is not a likely cause. Considering that the current dataset consisted of markers spread over 27 chromosomes, out of 38 present in the chicken genome, and that the estimates of genetic variance obtained using marker information did not detect all the genetic variance identified using BLUP PED, it appears that a proportion of "lost" variance could be explained by QTLs contained on chromosomes not represented in this dataset. These chromosomes, despite their short physical size, have been found to be gene rich (Hillier et al., 2004), thus it is highly likely that they contain genes of importance for these traits. This speculation is supported by the magnitude of the variance not captured in this study for BWT and LFI. The linkage length of the genome represented by markers used in the presented analyses accounts for around 80% of the length of the linkage map presented by Groenen et al. (2009). The proportion of the variance not captured by these markers for BWT and LFI shows remarkable agreement with this under-representation of the genome, at 27% for BWT and 20% for LFI. This would particularly affect the comparisons made with the observations from Daetwyler et al. (2008), as the equations presented were constructed on the assumption that markers capture all the variance present in a trait. Alternatively, as the analysed dataset is obtained from a working farm, it is possible that some nuisance factors have not been included in the models, leading to reduced ability of the model to identify genetic variance and consequently increasing error of predictions. However, this would affect both pedigree based and genomic predictions. It is also possible that the genetic structure of this particular

population could have contained some confounding effects, possibly attributable to family relationships. Finally, it is possible that due to the complex nature of the chicken genome, with highly variable recombination rates leading to variable lengths of the chicken genome between populations, the theoretical predictions presented in published studies and based on simulated mammalian data, may not be appropriate for estimation of performance of GP in chickens.

The effects of increasing marker density on estimates of variance components and predictive abilities of models diverged from the effects of increasing the TRN size. While increasing the number of markers used increased the estimate of the genetic variance, it had no clear effect on the accuracy of prediction. This discrepancy has been observed before (e.g. Makowsky et al., 2011, de los Campos, 2014). Since the early studies into the use of markers it has been expected that dense marker maps will be more efficient in capturing the genetic variance (Meuwissen et al., 2001, Hayes et al., 2007). This expectation is based on the infinitesimal theory, where the number of loci affecting the traits is large and thus requires dense marker maps. In broiler chickens, characterized by short stretches of LD as compared to layers (Andreescu et al., 2007), it has been suggested that as many as 100k markers would be necessary to capture the majority of the genetic variance (Megens et al., 2009). This high density of 100K markers was found to capture the highest proportion of variance in a study based of a small number (N=1,351) of broiler birds (Abdollahi-Arpanahi et al., 2014). Exceeding the 100K density did not bring further benefits in that study, with the proportion of the variance captured at 100K, 200K and 345K being similar, at 69 - 71% for BWT and 70 - 72% for HHP (Abdollahi-Arpanahi et al., 2014). In the presented results, increasing the density from 2K to 412K brought about an increase in the proportion of the genetic variance captured, however, the majority of the increase, particularly for BWT, was observed between chips up to 40K. It has been previously speculated that 10 - 20K markers with highest effect on the trait could suffice for routine evaluations of broilers, however the study failed to report what was the proportion of the genetic variance captured by such subset of markers (Abdollahi-Arpanahi et al., 2014). The results presented in this Chapter show that the required density can be lower than previously expected, and, with enough TRN records, a random sample of markers can be just as effective as pre-selected SNPs of highest effect on the trait.

The clear effect of increase in the proportion of the variance captured observed with increasing marker density is contrasted with lack of such improvement in predictive ability. The chip density at which the accuracy reaches plateau seems to be population specific, with

over 80K densities required in a population of largely unrelated humans (Makowsky et al., 2011) contrasted with 20K reported in layers (Wolc et al., 2011b) and 10K in broilers (Abdollahi-Arpanahi et al., 2014). In the latter, broiler study, the 10K chip was suggested as sufficient, as no changes in accuracy were observed for different marker densities across three traits, body weight, breast muscle and hen housed production. While the presented results corroborate with the 10K suggestion, with the accuracy largely unchanged when marker densities increase over 19K, a much larger dataset allowed detection of an increase in accuracy between chips 2K and 7K, not observed in the previous broiler study (Abdollahi-Arpanahi et al., 2014).

Makowsky et al. (2011) suggested that the differences in density at which the accuracy of prediction reaches a plateau are caused by different lengths of the LD blocks. However, remarkable agreement between layer and broiler studies suggests that some other factors affect the required marker density, as the stretches of LD in layers have been shown to far outreach those in broilers (Andreescu et al., 2007).

In a simulation study examining the effect of marker density on the accuracy of predictions, Solberg et al. (2008) stated that doubling the density of SNP markers increases accuracy of predictions by 1.04 - 1.07 fold. The results of this study show however that the accuracy approaches a plateau when marker densities exceed $4N_e$/Morgan. In the presented analysis, markers were spread over 27 chromosomes. Thus, a density of $27x4N_e$, at around 10.8K markers (N_e=100), becomes the lower limit for the required density. This stands in remarkable agreement with the empirical results for BWT and LFI accuracy, with accuracy improved slightly when density increased from 2K to 7K and no further significant changes observed with higher densities. The accuracy observed at these densities did not reach the 0.8 observed by Solberg et al. (2008), however, their approximation assumed a heritability of a trait equal to 0.5. Out of the three traits studied in this Chapter, only LFI had heritability estimated close to this value, however even this estimate is expected to be inflated, as the models fitted to it did not include other significant effects (see Chapters 1 and 2). Thus, due to lower heritability of the traits a lower accuracy was expected. Further, the results based on real data are bound to be burdened with more noise than simulated data, thus resulting in lowered predictive abilities.

The inconsistency of HHP results compared to a clear pattern observed in BWT and LFI can be explained by the number of individuals on which SNP effects were calculated being too low to capture enough information for accurate predictions. Low numbers of HHP records lead also to increased standard error estimates, which make precise comparison of method

across scenarios limited. Aside from formal estimates of standard errors, the fluctuation of the accuracy for this trait can be considered as an empirical error, which is caused by the complex structure of this trait and low numbers of records available. The distribution of HHP records has been shown to depart from normality (Ibe and Hill, 1988), which was an underlying assumption applied to all traits in this study. Treating HHP as a normally distributed trait results in increased heterogeneity of error variance between fixed effect levels, reducing the ability to correctly partition genetic variance and thus worsens the accuracy of BV prediction. Based on the results presented, the accuracy of genomic predictions for this trait can only be considered as a value in range 0.31 - 0.47.

While the accuracy estimates for HHP showed lack of consistency and counter-intuitive influence of the TRN size, the bias of predictions for this trait was the least out of the three traits analysed. However, the large standard errors for the regressions reflect the numbers of HHP records in the TST set, and lower heritability of this trait. When using the phenotype to assess accuracy, the lower heritability implies a greater environmental noise, with simultaneous lower variance of GEBVs. Therefore the regressions were well estimated for BWT and LFI, but had very large standard errors for HHP. The regressions for BWT and LFI indicate that the GBLUP model used in this analysis was likely overestimating differences in true breeding values.

Low bias of genomic predictions for BWT and LFI is an interesting observation, when compared to the considerable bias of pedigree-based predictions for these traits. Under BLUP methodology, the predictions of breeding values are expected to be unbiased, i.e. the expected regression coefficient should be 1 (Henderson, 1975). The departure from this value across methods suggests underlying selection bias, which was not accounted for in the model. This is not surprising, as the current selection process in broiler chickens is based on multiple traits. As such, predictions based on univariate analyses will not be able to correct for reduction of the variance due to selection on other traits.

This could also possibly explain the lower bias of genomic predictions. Considering that many of the traits used in genetic improvement programme are genetically correlated, genomic analyses on one trait may have been able to distinguish between animals which would rank as superior or inferior for another, correlated trait. In contrast, pedigree based analyses can only utilize the data provided directly in the model.

5. CONCLUSIONS

Analysis of a large number of chicken phenotypes and genotypes revealed that increasing the TRN population size is more consistent in increasing the accuracy of genomic prediction than increasing chip density. Comparison of accuracies obtained from different chip densities confirmed that increasing the number of markers over 20k does not bring additional benefits in accuracy. This finding stood in agreement with both empirical and simulated findings published in the literature. In contrast, the number of TRN individuals used in this study exceeded numbers defined previously in simulation studies, but failed to achieve speculated accuracies. Whilst all of the examined traits require more records to establish accuracies of 0.7 or more, the benefits of using marker information over traditional pedigree-based predictions are already considerable, with accuracy improved by as much as 77% for FI. This improvement is accuracy, aside from other possible benefits, will have a direct effect on gains obtained from broiler breeding programs. In addition, genomic methods are expected to better deal with selection bias than pedigree based methods.

6. REFERENCES

ABDOLLAHI-ARPANAHI, R., PAKDEL, A., NEJATI-JAVAREMI, A., MORADI SHAHRBABAK, M., MOROTA, G., VALENTE, B. D., KRANIS, A., ROSA, G. J. M. & GIANOLA, D. 2014. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. *Journal of Animal Breeding and Genetics,* 131**,** 183-193.

ABDOLLAHI-ARPANAHI, R., NEJATI-JAVAREMI, A., PAKDEL, A., MORADI-SHAHRBABAK, M., MOROTA, G., VALENTE, B., KRANIS, A., ROSA, G. & GIANOLA, D. 2014. Effect of allele frequencies, effect sizes and number of markers on prediction of quantitative traits in chickens. *Journal of Animal Breeding and Genetics*, 131, 123 - 133

ANDREESCU, C., AVENDANO, S., BROWN, S. R., HASSEN, A., LAMONT, S. J. & DEKKERS, J. C. 2007. Linkage disequilibrium in related breeding lines of chickens. *Genetics,* 177**,** 2161-9.

ANDREESCU, C., HABIER, D., FERNANDO, R., KRANIS, A., WATSON, K., AVENDANO, S. & DEKKERS, J. C. M. Accuracy of Genomic Predictions Across Breeding Lines of Chickens. World Congress on Genetics Applied to Livestock Production, 2010 Leipzig.

AVENDANO, S., WATSON, K. & KRANIS, A. Genomics in Poultry Breeding - From Utopias to Deliverables. World Congress on Genetics Applied to Livestock Production, 2010 Leipzig.

CHEN, C. Y., MISZTAL, I., AGUILAR, I., TSURUTA, S., MEUWISSEN, T. H. E., AGGREY, S. E., WING, T. & MUIR, W. M. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. *J Anim Sci,* 89**,** 23-28.

DAETWYLER, H. D., HICKEY, J. M., HENSHALL, J. M., DOMINIK, S., GREDLER, B., VAN DER WERF, J. H. J. & HAYES, B. J. 2010. Accuracy of estimated genomic

breeding values for wool and meat traits in a multi-breed sheep population. *Animal Production Science,* 50**,** 1004-1010.

DAETWYLER, H. D., VILLANUEVA, B. & WOOLLIAMS, J. A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One,* 3**,** e3395.

DE LOS CAMPOS, G., SORENSEN, D., GIANOLA, D. Genomic Heritability: What is it? 10th World Congress on Genetics Applied to Livestock Production, 18 Aug 2014 2014 Vancouver.

DE LOS CAMPOS, G., VAZQUEZ, A. I., FERNANDO, R., KLIMENTIDIS, Y. C. & SORENSEN, D. 2013. Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genet,* 9**,** e1003608.

ELFERINK, M., VAN AS, P., VEENENDAAL, T., CROOIJMANS, R. & GROENEN, M. 2010. Regional differences in recombination hotspots between two chicken populations. *BMC Genetics,* 11**,** 11.

GILMOUR, A. R., GOGEL, B. J., CULLIS, B. R., WELHAM, S. J. & THOMPSON, R. 2006. *{ASREML} User Guide Release 2.0*.

GODDARD, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica,* 136**,** 245-257.

GODDARD, M. E. & HAYES, B. 2007. Genomic Selection. *Journal of Animal Breeding and Genetics,* 124**,** 323 - 330.

GONZÁLEZ-RECIO, O., GIANOLA, D., ROSA, G. J., WEIGEL, K. A. & KRANIS, A. 2009. Genome-assisted prediction of a quantitative trait measured in parents and progeny: application to food conversion rate in chickens. *Genetics, Selection, Evolution* 41 (3).

GRAY, A., STEWART, I. & TENESA, A. 2012. Advanced complex trait analysis. *Bioinformatics,* 28**,** 3134-6.

GROENEN, M. A. M., WAHLBERG, P., FOGLIO, M., CHENG, H. H., MEGENS, H. J., CROOIJMANS, R. P. M. A., BESNIER, F., LATHROP, M., MUIR, W. M., WONG, K., GUT, I. & ANDERSSON, L. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research,* 19**,** 510-519.

HAYES, B. & GODDARD, M. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution,* 33**,** 209 - 229.

HAYES, B. J., CHAMBERLAIN, A. J., MCPARTLAN, H., MACLEOD, I., SETHURAMAN, L. & GODDARD, M. E. 2007. Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. *Genetical Research,* 89**,** 215-220.

HAYES, B. J., VISSCHER, P. M., MCPARTLAN, H. C. & GODDARD, M. E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res,* 13**,** 635-43.

HENDERSON, C. R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,* 31**,** 423-447.

HICKEY, J., GORJANC, G., KINGHORN, B., TIER, B., VAN DER WERF, J. H. & CLEVELAND, M. 2011. An imputation strategy which results in an alternative parameterization of the single stage genomic evaluation. *Interbull Bulletin 44*.

HILLIER, L. W., MILLER, W., BIRNEY, E., WARREN, W., HARDISON, R. C., PONTING, C. P., BORK, P., BURT, D. W., GROENEN, M. A. M., DELANY, M. E., DODGSON, J. B., CHINWALLA, A. T., CLIFTEN, P. F., CLIFTON, S. W., DELEHAUNTY, K. D., FRONICK, C., FULTON, R. S., GRAVES, T. A., KREMITZKI, C., LAYMAN, D., MAGRINI, V., MCPHERSON, J. D., MINER, T. L., MINX, P., NASH, W. E., NHAN, M. N., NELSON, J. O., ODDY, L. G., POHL, C. S., RANDALL-MAHER, J., SMITH, S. M., WALLIS, J. W., YANG, S. P., ROMANOV, M. N., RONDELLI, C. M., PATON, B., SMITH, J., MORRICE, D.,

DANIELS, L., TEMPEST, H. G., ROBERTSON, L., MASABANDA, J. S., GRIFFIN, D. K., VIGNAL, A., FILLON, V., JACOBBSON, L., KERJE, S., ANDERSSON, L., CROOIJMANS, R. P. M., AERTS, J., VAN DER POEL, J. J., ELLEGREN, H., CALDWELL, R. B., HUBBARD, S. J., GRAFHAM, D. V., KIERZEK, A. M., MCLAREN, S. R., OVERTON, I. M., ARAKAWA, H., BEATTIE, K. J., BEZZUBOV, Y., BOARDMAN, P. E., BONFIELD, J. K., CRONING, M. D. R., DAVIES, R. M., FRANCIS, M. D., HUMPHRAY, S. J., SCOTT, C. E., TAYLOR, R. G., TICKLE, C., BROWN, W. R. A., ROGERS, J., BUERSTEDDE, J. M., WILSON, S. A., STUBBS, L., OVCHARENKO, I., GORDON, L., LUCAS, S., MILLER, M. M., INOKO, H., SHIINA, T., KAUFMAN, J., SALOMONSEN, J., SKJOEDT, K., WONG, G. K. S., WANG, J., LIU, B., WANG, J., YU, J., YANG, H. M., NEFEDOV, M., KORIABINE, M., DEJONG, P. J., GOODSTADT, L., WEBBER, C., DICKENS, N. J., LETUNIC, I., SUYAMA, M., TORRENTS, D., VON MERING, C., et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature,* 432**,** 695-716.

IBE, S. & HILL, W. 1988. Transformation of poultry egg production data to improve normality, homoscedasticity and linearity of genotypic regression. *Journal of Animal Breeding and Genetics,* 105**,** 231-240.

KRANIS, A., GHEYAS, A., BOSCHIERO, C., TURNER, F., YU, L., SMITH, S., TALBOT, R., PIRANI, A., BREW, F., KAISER, P., HOCKING, P., FIFE, M., SALMON, N., FULTON, J., STROM, T., HABERER, G., WEIGEND, S., PREISINGER, R., GHOLAMI, M., QANBARI, S., SIMIANER, H., WATSON, K., WOOLLIAMS, J. & BURT, D. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics,* 14**,** 59.

LUAN, T., WOOLLIAMS, J. A., LIEN, S., KENT, M., SVENDSEN, M. & MEUWISSEN, T. H. E. 2009. The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation. *Genetics,* 183**,** 1119-1126.

MAKOWSKY, R., PAJEWSKI, N. M., KLIMENTIDIS, Y. C., VAZQUEZ, A. I., DUARTE, C. W., ALLISON, D. B. & DE LOS CAMPOS, G. 2011. Beyond missing heritability: prediction of complex traits. *PLoS genetics,* 7**,** e1002051.

MEGENS, H.-J., CROOIJMANS, R., BASTIAANSEN, J., KERSTENS, H., COSTER, A., JALVING, R., VEREIJKEN, A., SILVA, P., MUIR, W., CHENG, H., HANOTTE, O. & GROENEN, M. 2009. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. *BMC Genetics,* 10**,** 86.

MEUWISSEN, T. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution,* 41**,** 35.

MEUWISSEN, T., HAYES, B. & GODDARD, M. E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics,* 157**,** 1819 - 1829.

PRYCE, J. E. & DAETWYLER, H. D. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Animal Production Science,* 52**,** 107-114.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. W., DALY, M. J. & SHAM, P. C. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics,* 81**,** 559-575.

RODIONOV, A. V. 1996. Micro vs Macro: A review of structure and functions of avian micro- and macrochromosomes. *Genetika,* 32**,** 597-608.

SIMEONE, R., MISZTAL, I., AGUILAR, I. & VITEZICA, Z. G. 2012. Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure. *Journal of Animal Breeding and Genetics,* 129**,** 3-10.

SOLBERG, T., SONESSON, A. & WOOLLIAMS, J. 2008. Genomic selection using different marker types and densities. *Journal of Animal Science,* 86**,** 2447-2454.

WOLC, A., ARANGO, J., SETTAR, P., FULTON, J., O'SULLIVAN, N., PREISINGER, R., HABIER, D., FERNANDO, R., GARRICK, D. & DEKKERS, J. 2011a. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genetics Selection Evolution,* 43**,** 23.

WOLC, A., STRICKER, C., ARANGO, J., SETTAR, P., FULTON, E., O'SULLIVAN, N., HABIER, D., FERNANDO, R., GARRICK, D., LAMONT, S. & DEKKERS, J. C. M. Breeding Value Prediction For Production Traits In Layers Using Pedigree and Marker Based Methods. World Congress on Genetics Applied to Livestock Production, 2010.

WOLC, A., STRICKER, C., ARANGO, J., SETTAR, P., FULTON, J., O'SULLIVAN, N., PREISINGER, R., HABIER, D., FERNANDO, R., GARRICK, D., LAMONT, S. & DEKKERS, J. 2011b. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution,* 43**,** 5.

Chapter 7

GENERAL DISCUSSION

The analyses presented in this thesis provide an insight into the challenges of using genomic prediction in broiler genetic improvement programs and give an update on the classical parameters used in traditional evaluations. Vast amounts of data, in form of pedigrees, phenotypic records and genotypes, as well as novel techniques, such as regression of genomic relationships back to pedigree with the optimum regression coefficient estimated from the data, allowed an extensive evaluation of the population and methodologies.

1. THESIS OVERVIEW

Compared to other livestock species, the routine use of genomic selection in broiler chickens is still in its infancy. The commercial broiler evaluation procedures have, until recently, been based on BLUP methodology using phenotypes and pedigree only (Wolc, 2014), where the covariances between breeding values of selection candidates are calculated from the pedigree based relationships between these individuals and the estimates of the genetic variance present in a trait in question (Henderson, 1975). While the broiler companies are likely to carry out regular re-estimation of the variance components, most of the published estimates originate from 1990's (e.g. Chambers, 1990, Koerhuis et al., 1997, Danbaro et al., 1995), thus creating a gap in the publicly available estimates of the heritability for broiler traits. These estimates are necessary for the evaluation of the genomic analyses on the data, as they allow identification of the proportion of the variance captured by the markers. Chapter 2 of this thesis provides the current estimates of the variance components for 6 broiler traits, representing growth, feed efficiency and reproduction. These traits were analysed using several models, differing in the choice of random effects fitted. The results showed an interesting presence of significant maternal genetic and environmental effects for most of the traits studied, including traits recorded late in life of a bird. Chapter 3 continued the BLUP pedigree based analyses, with a vast data size of 1.3M records spread over 24 generations, which allowed the evaluation of the changes in the estimates of the genetic variance over time. It was shown that the genetic variance of juvenile body weight (BWT) has been decreasing over the past decade, while the genetic variance of hen housed production (HHP)

remained relatively constant. This Chapter provided also previously uncommon estimates of the accuracy and bias of pedigree based predictions for these two traits.

While the BLUP analyses based on pedigree have been an important part of the genetic improvement programme, the attention of breeders is now being turned toward genome-wide evaluations, which utilize marker information to calculate the GEBVs. Chapter 4 presents the preliminary results of the genomic evaluation in broilers, with the use of genotypes of up to 600K SNPs. The analyses in this Chapter concentrated on the traits which are expected to benefit most from the use of marker data: sex-limited reproductive traits, and difficult to measure fitness such as of overall mortality. The analyses included also moderately heritable key broiler traits, BWT and HHP. The anticipated rise in the accuracy of predictions with the use of genomic relationship matrices was proved for some traits, particularly for early hatchability (EHOF), while for other traits the desired effect was not observed in the analysed dataset. Chapter 6 showed that this lack of improvement in accuracies observed in Chapter 4 was most likely caused by insufficient numbers of phenotypic records. The analyses in Chapter 6 were based on a population size far exceeding the broiler populations previously studied in genomic selection experiments, with numbers of individuals with both phenotypes and genotypes reaching over 23.5K. This large number of records allowed a more in-depth analysis of the effect of the TRN size on the accuracy of predictions and showed the discrepancy between empirical accuracy and accuracy predicted based on population parameters. The latter expectations of accuracy were obtained from theoretically derived formulae, validated in simulation studies (Meuwissen, 2009, Daetwyler et al., 2008). Both Chapters 4 and 6 showed that increasing marker density brings limited improvement to the prediction accuracy. Chapter 5 compared the different use of marker data in calculation of genomic relationship matrices, with the matrices fitted in the analyses based on an LD or LA approach, and a mixture of both. It was shown that the performance of the (G)EBV prediction depends on the methodology used and that the best choice of the relationship matrix varies between different subsets of markers, even at the same marker density.

2. Up to Date Estimates of Variance

Despite the prolonged, intensive and highly specialized selection, the amount of the genetic variance present in the broiler line studied is still at a relatively high level for most traits, with the estimates of the heritability obtained from REML based on pedigree ranging between 0.04 (mortality) and 0.79 (egg weight), depending on the trait, choice of model terms and amounts of data. The heritability of the key broiler trait, BWT, estimated on the

data prior to 2011 (datasets analysed in Chapters 2 and 3) showed a surprising agreement with the estimates obtained from studies published two or more decades ago (as reviewed by Chambers, 1990), with heritability varying between 0.3 and 0.5. This maintained estimate of the heritability in Chapters 2 and 3 is surprising due to the large selection pressure applied to this trait, illustrated by the changes of the estimate of the genetic variance over time presented in Chapter 3. The similarity between the estimates of the genetic variance presented in this thesis and in previous reports is even more surprising, considering the differences in models used to obtain them. The estimates presented in Chapters 2 and 3 were obtained from the models characterized by the best fit to the data, which included maternal genetic and environmental effects. In contrast, most of the published studies were based on models which did not include maternal genetic effects. Omission of significant maternal effects results in inflated estimates of the genetic variance, as was shown in Chapters 2 and 3, and other studies into the effect of model choice (e.g. Clément et al., 2001).

The pedigree-based estimates of heritability for the genotyped populations obtained in Chapters 4 - 6 centre around 0.3. Considering the inflation of the genetic variance estimates when maternal effects are missing from the model for BWT, it appears that the most recent generation of broilers in this line (represented by the genotyped population) show slightly reduced genetic variance, with the estimates of heritability being lower than 0.3

Furthermore, the estimates of the genetic variance for BWT obtained in this thesis are likely to be overestimating the true variance for this trait, as indicated by the estimates of bias for the predictions of (G)EBVs for this trait, consistently lower than 1 across the analyses. While BLUP methodology is expected to account for the effect of selection on the variance in an offspring population if all selective criteria are included in the data (Henderson, 1975), the modern selection programs in broilers are based on multiple traits, with variable degrees of covariance between the component traits. Thus, estimation based on univariate analyses, albeit including BWT records for all selection candidates, is not likely to correct for the selection acting on all traits. Interestingly, in most cases the bias of genomic predictions was less than that of the pedigree based predictions, perhaps due to the direct information on inheritance of the LD blocks sourced from the genotypes.

3. MISSING VARIANCE

The lower bias of the genomic predictions is somewhat surprising, as the proportion of the genetic variance captured by markers was less than 1 for nearly all analyses presented in this

study. This observation of "missing variance" is fairly common across species. The idea of "missing heritability" was first realized when summation of the effects of GWAS hits for human height, unsurprisingly, accounted for a very low proportion of the phenotypic variance (Visscher, 2008). Soon after, it was realized that fitting multiple markers simultaneously explains a larger proportion of variance than the QTLs individually but still does not capture all the genetic variance identified through the pedigree analyses (Yang et al., 2010). The typical explanations for this phenomenon include the incomplete linkage between markers and QTLs (de los Campos et al., 2013), data structure and unrepresented parts of the genome. The latter seems to be particularly pertinent to poultry, where only 27 out of 38 autosomes are represented in most of the studies (e.g. Groenen et al., 2009, Gheyas and Burt, 2013). Thus, several microchromosomes, some with known QTLs and all with reported high gene-density, are usually omitted from genome-wide evaluations, due to the poor quality of the signal.

Considering the lowered proportion of variance captured by the markers, the interpretation of the heritability obtained from the marker data thus differs from the interpretation of the heritability obtained from the classical inifinitesimal model applied to pedigree based analyses only (de los Campos, 2014). It is assumed, that due to finite numbers of markers used and the imperfect linkage disequlibrium between markers and QTLs, the proportion of the variance they capture will inevitably be lower than a 100% (de los Campos et al., 2013). However, it would be interesting to evaluate if this observation holds for traits with low phenotype expression particularly when their heritability is low, e.g. mortality, for which the detection of the genetic variance using pedigree based methods is already difficult. The pedigree based estimates of the genetic variance for mortality in broiler chickens are notoriously low, as was shown in Chapter 4 and other published studies (e.g. Long et al., 2007). This results in poor response to selection, with mortality rates remaining largely unchanged (Besbes and Ducrocq, 2003). However, this trait can be successfully selected upon when the associative effects between individuals are included in the models, as shown by a reduction in mortality rates from 68% to 8.8% achieved through group selection carried out over 6 generations (Muir, 1996). The models fitted to mortality (MORT) records in Chapter 4 showed that markers, particularly on higher density chips, estimated the genetic variance at a higher value than the analysis based on the pedigree. Large standard errors brought about by the low number of records for this trait make the formal comparison impossible, however, this observation has been reported before in a study into the genomic selection applied to broiler mortality (Gonzalez-Recio et al., 2008). There are several possible explanations for this observation, ranging from a random occurrence of these

estimates within the standard error, through higher sensitivity of marker data than can be achieved for this trait with pedigree analyses, to upward bias of marker estimates, brought about by a large number of markers in linkage equilibrium with QTLs affecting the trait (de los Campos, 2014). Thus, a larger study with more MORT records would be an interesting experiment. Reducing the standard error of the estimates would hopefully allow more informed interpretation of the results.

The incomplete LD between markers and QTLs is likely the most frequently speculated cause of the missing variance in marker based inference of variance components (de los Campos et al., 2013). This explanation has been developed from the somewhat naïve assumption that high marker densities ensure that each QTL is in linkage disequilibrium with at least one nearby marker (Calus, 2010, Goddard et al., 2010), which was also supported by studies showing increases in accuracy of GEBV prediction with increased marker densities (e.g. Solberg et al., 2008). The results presented in Chapters 4 and 6 of this thesis illustrate however that increasing marker density over 40K by random addition of SNPs may not bring desired improvement in the detection of genetic variance. In the analysed data, increasing marker density resulted in a minor increase in the amount of variance captured, particularly for high density chips. For example, increasing marker density for BWT from 2K to 40K in Chapter 6 resulted in an increase of the variance captured from 52% to 68% (calculated as a proportion of the pedigree-based estimate of the genetic variance), while increasing it further from 40K to 412K resulted in only 5% additional gain. Similarly for HHP, the proportions of variance captured by 2K, 40K and 412K chips were 42%, 52% and 57% respectively.

Further, the impact of imperfect LD between markers and QTLs is likely to be reduced in poultry, where the populations studied usually include large numbers of relatives, with large full- and half-sib families. A simulation study of human data showed that estimations carried out in such populations are less sensitive to the imperfect LD than populations consisting of nominally unrelated individuals (de los Campos et al., 2013).

The proportion of variance captured depended more on the size of the data, population used and fixed effects fitted in the model. The proportion of the genetic variance for BWT captured in Chapter 4, between 75 - 94%, was considerably higher than the proportion captured in Chapter 6, between 52 - 73%. This can be in part attributed to the different choice of fixed effects used in the models in the two Chapters, with simple effects of hatch week and sex in Chapter 4 contrasted with the comprehensive, combined factor in Chapter 6. The simplified effects fitted in Chapter 4 resulted in the PB estimate of the genetic variance for this trait being nearly 10% lower than the equivalent estimate in Chapter 6. In contrast,

the estimates of the genetic variance obtained through REML using **G** matrix in Chapter 4 were higher than the equivalent estimates in Chapter 6. It is possible that the REML analyses in Chapter 4 assigned to the additive effects of the selection candidates some genetic differences between individuals in contemporary groups, which were accounted for in the large factor used in Chapter 6. An obvious factor choice would be the "mating group" effect, which codes for the hatch week of the parents of a selection candidate, thus possibly accounting for genetic trends.

For HHP, the proportion of the variance explained between the Chapters varied as well, between 64 - 90% captured in Chapter 4 and 42 - 57% in Chapter 6. For this trait, the two populations varied not only in the numbers of records available, but also in the choice of the genotyped birds. The dataset analysed in Chapter 6 included a group of selection rejects, thus widening the total variance observed with estimates of both genetic and error variances increased, while analyses presented in Chapter 4 were based on a small group of the candidates that passed several stages of selection, thus representing a narrowed distribution of the phenotypes for this trait.

The proportion of the variance captured by genomic data is also likely to depend on the choice, rather than density, of the markers used. An analysis of two chips with the same marker density but different choice of markers in Chapter 5 showed the impact of the marker location in the genome, with markers selected through the GWAS procedure capturing more of the variance than evenly spaced markers. Thus, an examination of marker location, as well as density, should be considered in relation to the proportion of the variance captured by these markers.

Another speculation regarding the missing variance is based on MAF screening criteria, with a proportion of the low frequency QTL alleles suspected to be lost through removal of markers with low MAF (Daetwyler, 2009). This abundance of QTL associations for low MAF markers has been confirmed in a study of juvenile body weight in broilers, where markers with MAF<0.2 explained 75% of the variance explained by markers (Abdollahi-Arpanahi et al., 2014). This result seems somewhat questionable, as, considering a relatively small sample of that study (n=1,351), the frequencies reported could be an artefact of a small population size. Nevertheless, the proportion of markers removed due to low MAF (with a threshold set at 0.01) in this thesis was considerable, at 25% of the full range of 600K markers removed from the data. Thus, it is possible that some rare QTLs were indeed lost.

4. ACCURACY OF GENOMIC PREDICTIONS

Traditionally, the accuracy of genomic prediction has been considered as a function of the effective population size, number of independent chromosomal segments segregating in the population, heritability of a trait and number of phenotyped and genotyped individuals (Daetwyler et al., 2008). Several deterministic predictions of achievable accuracy exist. In Chapter 6, these deterministic predictions were compared with the empirical results based on the real broiler data. A considerable discrepancy has been found between the accuracy observed, and those expected given the trait and population characteristics. However, the accuracy obtained using genomic predictions is limited by the proportion of the genetic variance that is not captured by the markers. This proportion varied between traits and population sizes in the presented results, as discussed before. Daetwyler (2009) showed that empirical accuracy can be used to determine the proportion of the genetic variance captured in the particular dataset (q^2), as:

$$q^2 = \frac{r^2(Nc+1)}{Nc}$$

where N is the number of TRN records, r^2 is the reliability of the prediction calculated as the square of the accuracy of GEBV prediction and $c=h^2/M_e$.

Thus, substituting the values describing the BWT data from Chapters 4 and 6, and assuming M_e=771, the proportion of variance that was captured for this trait, estimated from the observed accuracies yields:

- 20% for Chapter 3, where N=3,162, h^2=0.35 and $r^2=(0.34)^2$
- 30% for Chapter 5, where N=21,225, h^2=0.41 and $r^2=(0.53)^2$

The q^2 estimated from observed accuracies is thus considerably lower than the proportion of the variance captured by markers, calculated as a ratio between the genomic estimate of the genetic variance and a pedigree-based reference. The large discrepancy between the two methods illustrates the separation between the inference of variance components and the predictive abilities, which appears to be more dependent on the number of records available in TRN rather than on the estimate of the genetic variance per se. The discrepancy between the predictive ability, and variance captured by the same set of markers has been shown before (Daetwyler et al., 2013).

The theoretically derived formulae estimating the accuracy of genomic predictions frequently utilize the M_e parameter, which partitions the genome into independent segments (e.g. Daetwyler et al., 2008, Goddard, 2009). However, this term appears to be difficult to estimate in practical applications, with different values obtained depending on the method of calculation. For example, the formulae for calculating the M_e proposed by Goddard (2009) and further used by others, e.g. Daetwyler et al. (2008), is based on the effective population size and the length of the genome in the form of $\frac{2N_eL}{\log(4N_eL)}$. In contrast, Hayes et al. (2009) estimates the number of independent segments from Mendelian sampling variation between family members, or, for a random mating population uses a simplified formula of $2N_eL$. This parameter has been also used in calculation of the regression coefficients used for shrinkage of the genomic relationships back to the values expected from the pedigree (Goddard et al., 2011), as shown in Chapter 5. Substituting the chicken genome parameters to Goddard's formula gives a wide range of estimates, between M_e=294 (N_e=50 and L=25M) and M_e=1,438 (N_e=200 and L=37M). Further, using the regression coefficients empirically identified as best fitting the data in Chapter 5 give extreme values of M_e: up to 108,000 for the ESM chip and 27,000 for GWAM chip. Thus, while this parameter may be a useful tool for theoretical derivations and simulations, its applicability to real data is controversial.

5. MATERNAL MODELS

In the absence of genomic information, the best predictions in terms of bias were obtained by fitting models which included maternal effects. It is possible that these effects accounted for some of the pre-selection practices not included in the available datasets and thus reducing the ability of BLUP PED to detect the reduction of variance due to the truncated distribution of records. Models including maternal effects were also shown to fit the data with highest likelihood (as shown in Chapter 2 and 3). Thus, inclusion of maternal effects in models can be considered as beneficial for most traits, as it:

- improves the fit of the models to data
- improves the partitioning of the variance, resulting in more accurate estimates of the genetic variance and heritability
- reduces bias of predictions
- provides a new source of variation which can be selected for.

While the magnitude of maternal effects seems to be unaffected by the amount of data, the precise estimation of the correlation between maternal genetic and direct effects requires extremely large numbers of records, with the estimates for BWT ranging from -0.02 (SE 0.06) for 590K individuals in Chapter 2 to 0.11 (SE 0.04) for 1.3M chickens in Chapter 3. Due to the low magnitude of the estimate and relatively large standard error, it is impossible to state which value represents the true correlation between these effects in BWT, however, the model resulting in the positive estimate in Chapter 3 showed a better fit to the data, with the AcM model characterized by the highest likelihood. A particular consideration needs to be given to traits for which this correlation is consistently negative, e.g. EWT, where selection on the direct effects will result in correlated negative effects in the maternal variation, thus possibly reducing the observable response to selection.

It is possible that the use of genotypes would help reduce the required number of records for estimation of the maternal effects. GP is expected to improve the evaluations for sex limited traits, as in contrast to traditional pedigree based estimation, it can source the information for each individual, directly from their genotype. Maternal genetic effects can be treated as a composite trait, incorporating a range of observed and unobserved characteristics of the dam, affecting her offspring. Although selection on each of the measurable components of the maternal influence has been previously suggested (Koerhuis and McKay, 1996), this has not been introduced into standard practice, partly due to high costs of phenotypic measurements collected on one sex only at the late stages of life, and partly due to low heritability of many of such traits (Dunn, 2011). Genomic prediction is expected to improve predictions of such traits, as it will be able to source information on both sexes through the use of the genotypes, with the possible additional benefit of shortening the generation interval. However, as was shown in Chapter 4 and 6, the number of phenotypes required for accurate estimation of marker effects is large, particularly for traits with low heritability. Thus, selection on particular components of the maternal effects would still pose a large monetary challenge in the collection of the phenotypes for each of these traits. In contrast, genomic models including maternal effects fitted to traits such as BWT could optimize the data usage. The maternal GEBVs could be extracted from already collected BWT phenotypes, and through the use of genotypes could be calculated with high accuracy for both males and females in the data.

6. REQUIRED DATA SIZE

The analyses presented in this study benefited from large amounts of data, with phenotypes recorded for up to 1.3M birds and over 23.5K birds' genotypes collected using the 600K Affymetrix Axiom chip. This wealth of data was attainable due to the structure of the breeding programs in broilers, where the superior animals in a generation are selected from a large population of selection candidates with phenotypic records, with quick accumulation of the data due to short generation intervals. The 1.3M data used in Chapter 3 were collected over a relatively short period of 15 years. Considering the large effect of the TRN size on the accuracy of genomic predictions, this gives high hopes for the future of GS in broilers, with the numbers of animals potentially available for genotyping reaching nearly 1M within just 10 years. However, the optimum mode of increasing the TRN size (i.e. increase in the number of individuals genotyped per generation, vs. accumulation of the data over generations) remains to be determined. Studies based on simulations (Muir, 2007) and real data (Wolc et al., 2011) showed that increasing the number of generations in TRN improves the accuracy of genomic predictions. This benefit of increased number of TRN generations is speculated to exceed the benefits of increasing the number of individuals within the generation (Muir, 2007). However, neither of the two studies considered a situation where the numbers of individuals would increase by the same numbers but within, rather than across, generations. As such, the observed increase in the accuracy of predictions could be easily attributed to rising number of TRN individuals overall, rather than through the increasing number of generations in TRN. Further, in multi-generation datasets the effects of selection need to be taken into account. Considering the reduction of the genetic variance over generations illustrated in Chapter 3, which was not accounted for by the BLUP methodology, and reduced bias of genomic predictions in Chapters 4 - 6, an analysis of changes of variance over time using genomic data would provide more information on the preferred method of data accumulation.

While the requirement for large numbers of TRN individuals, confirmed on a large dataset in Chapter 6, will pose a financial and logistic challenge to the routine use of GS in broilers, the findings of this study regarding marker density provide a potential way to limit the costs. It appears that increasing marker density over 40K in smaller datasets (Chapter 4) and 20K when large TRN is available (Chapter 6) brings limited improvement in the prediction accuracy. This is consistent with the conclusion that the extent of LD is relatively large in the studied line of broiler chickens. The 20-40K density at which the accuracy plateaued in the presented analyses follows the conceptual expectations of the intermediate optimum,

between low densities facing possible loss of variance captured, and high densities requiring a large financial input and possibly introducing confounding effects (Muir, 2007).

This, combined with a high accuracy of the imputations from low density to high density panels (Habier et al., 2009) implemented in programs such as Alpha Impute (Hickey et al., 2011), suggests that the costs of genotyping could be substantially reduced (Wolc et al., 2014).

The benefits of GP shown in Chapters 4 and 6 of this thesis were based on chips created from randomly selected markers, which form a universal base for multi-trait evaluations. In Chapter 5 of this thesis, and in some previously published papers (e.g. Calus et al., 2008), chips consisting of evenly spaced markers presented benefits over chips selected randomly, or based on GWAS hits. The lack of improved accuracy of the GWAM chip in Chapter 5 stands in contrast with some other published studies on GP in broilers, where pre-selection of markers based on their estimated effect on the trait has been presented as a preferred method, showing the optimum accuracy of predictions (e.g. Gonzalez-Recio et al., 2008, Long et al., 2007, Abdollahi-Arpanahi et al., 2014). However, such an approach requires prior estimation of marker effects for each trait. Further, to accurately estimate these marker effects and avoid ascertainment bias, large numbers of phenotypes are required, collected on animals genotyped for all the markers. Thus, while the evaluations of selection candidates are carried out with the reduced costs of genotyping for a subsample of selected markers, the costs incurred in estimation of the marker effects in training will face all the challenges of high density genotyping. Wolc et al. (2011) showed that each generation separating the selection candidates from the training set decreases the accuracy of predictions and based on this result recommended re-training in every generation. It follows that the impact of particular markers on the trait is likely to change as a result of selection, due to changed marker allele frequency and selection-induced linkage, therefore, pre-selection of markers with highest effect on the trait calls for frequent re-estimations in selected populations. All these procedures would have to be carried out for every trait included in the selection index, with a possibility of a different set of markers identified for each trait. Furthermore, de los Campos et al. (2013) showed that some of the methods of marker pre-selection, specifically ranking based on p-value obtained from GWAS analysis, has reduced impact in populations which contain related individuals, thus casting doubts on the use of such methods in livestock, particularly so in broiler populations characterized by large full- and half-sib families, with a multitude of other relationships. Therefore while pre-selection of markers is conceptually appealing, it may pose more challenges than benefits in the routine evaluations of broilers.

While the results of this study indicate that the available marker densities are sufficient for accurate estimation, the evaluations could possibly benefit from a better genome coverage. As discussed above, it is possible that a proportion of the genetic variance not captured by the markers used in this dataset could be explained by QTLs located on microchromosomes, not represented in this or other studies into GP in broilers. While the creation of the 600K Affymetrix Axiom Chip was based on sequence data of nearly 250 chickens, the assignment of the 139M SNPs to localized position in the genome was incomplete, due to gaps in the current chicken genome map. Thus, while increasing marker density over intermediate values is not likely to increase the accuracy of predictions, as shown in this thesis, a more complete genome coverage at a constant marker density could possibly improve the overall performance of the genomic evaluations.

7. WHAT NEXT?

The results of the analyses presented in this study, aside from providing up to date estimates of variance components and the requirements of the successful application of GP in broilers, provide also a framework for future research.

The main challenge for the future use of GP in broilers will be based on computational power, software limitations and data storage. Considering the progress in the performance of computing hardware is doubling every two-years, as predicted by Moore's Law (Moore, 1965), the computational power will soon be able to deal with large quantities of data generated by genotyping experiments on livestock. At the same time, multiple methodologies are being developed, with some aimed at optimization of the computational resources, e.g. software for genetic analyses based on parallel computing (e.g. Gray et al., 2012, Wang et al., 2014), while others modify the statistical approach to genetic analyses making them computationally less demanding and/or faster (e.g. Gorjanc et al., 2014).

While the largest analysis of this study, based on 23.5K genotypes containing 412K markers, took an acceptable time of nearly 49 hours overall (19h 20mins for the calculation of the genomic relationship matrix, 20h 30mins for REML analysis, 8h 43mins for estimation of marker effects and 24mins for prediction of GEBVs from the marker effects and genotypes), at the time of the analysis it was impossible to run a GBLUP multi- or even bi-variate analysis for such numbers of birds. Considering that the selection of broiler chickens is based on multiple traits, most of which are moderately correlated (as shown for several traits in Chapter 2), extending the analyses to multivariate methods will be necessary.

Another area of future research should concentrate on the genomic evaluations of the maternal effects. The use of marker data on both females and males may result in more precise estimates of these effects than can obtained from pedigree and dam records only. This could shed more light on the mechanism of maternal influence on the broiler traits and answer some of the inconsistencies found between different studies, e.g. the sign and magnitude of the correlation between maternal and direct genetic effects in BWT. The models used in the genomic analyses presented in Chapters 4 - 6 were limited to one random effect of the chick on which the data was collected, in contrast to widely accepted models including the permanent environment of the dam fitted to most traits. While the model choice is not expected to change the accuracy of predictions, repetition of these analyses with the equivalent model containing at least the permanent environment effect, could provide a more accurate inference of the variance components, as was shown for pedigree based predictions in Chapters 2 and 3. Similarly, fitting maternal effects is not expected to increase the accuracy of direct GEBV prediction, however, it may open up a previously untapped source of genetic variance. Considering the reduction of the genetic variance brought about by selection, such additional sources can prove valuable.

While the analyses presented here covered many of the key broiler traits, the numbers of records available for reproductive traits was too low to establish the potential benefits this method could bring in the future. It is expected that the progress achievable through GP applied to sex-limited traits will far exceed the progress currently provided by the pedigree-base estimations based on phenotypic records of one sex only. However, considering the low heritabilities of these traits, far larger numbers of records than were used in this study will be necessary.

Table 7.1 Required numbers of TRN records for accuracy of 0.9, calculated using formulas derived from Meuwissen (2009) in row 1, and Daetwyler et al. (2008) in row 2. The full description of this calculation was given in Chapter 5. Traits used are early (EHOF) and late (LHOF) hatchability, early (EFERT) and late (LFERT) feritility, and mortality (MORT). The heritabilities of these traits calculated based on BLUP PED are given in brackets. The genome parameters used are: effective population size N_e=100, genome length L=37M, number of independent chromosomal segments segregating in the population M_e=771.

Method	EHOF (h^2=0.21)	LHOF (h^2=0.11)	EFERT (h^2=0.10)	LFERT (h^2=0.11)	MORT (h^2=0.04)
$\frac{1.6N_eL}{h^2}$	28,190	53,818	59,200	53,818	148,000
$\frac{0.81M_e}{0.19h^2}$	15,652	29,904	32,895	29,904	82,237

Table 7.1 shows the numbers of TRN records required for these traits based on the formulae derived from simulation studies (Daetwyler et al., 2008, Meuwissen, 2009) and discussed in depth in Chapter 6. It is clear that those numbers exceed the number of currently genotyped and phenotyped birds. Moreover, the realistic expectations based on empirical proofs provided in this thesis, and a comparison of the numbers obtained from these formulae to observed results, indicate that these large numbers are still only a conservative underestimate of the true numbers needed. Albeit these numbers are large, they could be conceivably met considering the structure of the breeding programs in chickens, as was outlined before.

Alternatively, methods of increasing the numbers of the available phenotypes for the estimation of marker effects, outside of the pool of genotyped birds, could provide more instant returns than waiting for the records of genotyped birds to accumulate. Some authors have suggested that in such situations, a combination of the genomic relationship matrix with pedigree based relationships calculated for individuals without genotype, could provide a cost effective method of using all of the available data (Legarra et al., 2011). Perhaps the most widely studied approach to this combination has been grouped under "single step" term, where the relationship matrix **H** combines both genotyped and ungenotyped individuals, with the relationships of the latter corrected for the differences observed between pedigree and genomic relationships calculated within the genotyped group (Forni et al., 2011). However, this approach appears to work best when non-genotyped individuals are offspring to the genotyped animals (Meuwissen, 2014). An alternative approach is based on the extension of the genomic relationships back through the pedigree to the non-genotyped ancestors, through the linkage analysis approach which has been discussed in Chapter 5. The analyses based on LA and LDLA approach presented in this Chapter were limited both in terms of the number of markers and numbers of individuals used. Thus, the further evaluation of the methods with a larger dataset would be of interest. The analyses presented were limited to genotyped individuals only, however it would be interesting to compare the efficacy of the LA and single step methods in combining the additional information from ungenotyped animals with the data on genotyped birds. Further, as the regression coefficients used in the LDLA approach presented in Chapter 5 were dependent on the choice of markers, it would be interesting to evaluate the choice of optimal coefficients as the size of the TRN population and depth of the pedigree increase.

The obvious next step in the GP experiments in broilers is the validation of the presented results in other broiler, and possibly layer, lines. The performance of GP is likely to show similar patterns in the effect of TRN size and marker density to the ones shown here,

however, their extent will depend on the structure of the population, particularly on its LD block structure, which has been shown to be highly variable between broiler and layer lines (Aerts et al., 2007, Andreescu et al., 2007). Another parameter of the population under study which could be examined is the degree of relationships present between TRN and TST samples. Samples consisting of related individuals have been shown to be more immune to the impact of imperfect LD than un-related contemporaries (de los Campos et al., 2013) and to provide higher accuracy of predictions than unrelated individuals (Clark et al., 2012). However, the benefits of improved estimates on relatives need to be weighed against the risks of increased inbreeding.

8. CONCLUSIONS

In conclusion, the analyses presented in this thesis show a promising future for the use of Genomic Prediction in broiler improvement programs. This future of course will face some challenges, related mainly to the vast amounts of data required, however, thanks to the structure of broiler populations, this wealth of data will allow realization of the desired improvements in sex-limited traits, e.g HHP, while simultaneously allowing consideration of aspects, such as maternal genetic effects, which was not previously attempted in poultry.

9. REFERENCES

ABDOLLAHI-ARPANAHI, R., PAKDEL, A., NEJATI-JAVAREMI, A., MORADI SHAHRBABAK, M., MOROTA, G., VALENTE, B. D., KRANIS, A., ROSA, G. J. M. & GIANOLA, D. 2014. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. *Journal of Animal Breeding and Genetics,* 131**,** 183-193.

ABDOLLAHI-ARPANAHI, R., NEJATI-JAVAREMI, A., PAKDEL, A., MORADI-SHAHRBABAK, M., MOROTA, G., VALENTE, B., KRANIS, A., ROSA, G. & GIANOLA, D. 2014. Effect of allele frequencies, effect sizes and number of markers on prediction of quantitative traits in chickens. *Journal of Animal Breeding and Genetics,* 131, 123 - 133

AERTS, J., MEGENS, H. J., VEENENDAAL, T., OVCHARENKO, I., CROOIJMANS, R., GORDON, L., STUBBS, L. & GROENEN, M. 2007. Extent of linkage disequilibrium in chicken. *Cytogenet Genome Res,* 117**,** 338-45.

ANDREESCU, C., AVENDANO, S., BROWN, S. R., HASSEN, A., LAMONT, S. J. & DEKKERS, J. C. 2007. Linkage disequilibrium in related breeding lines of chickens. *Genetics,* 177**,** 2161-9.

BESBES, B. & DUCROCQ, V. 2003. Use of mixed model methodology in breeding strategies for layers. *Poultry Breeding and Biotechnology. W. Muir, and S. Aggrey, ed. CABI Publishing, Wallingford, UK***,** 127-146.

CALUS, M. P. L. 2010. Genomic breeding value prediction: methods and procedures. *Animal,* 4**,** 157-164.

CALUS, M. P. L., MEUWISSEN, T. H. E., DE ROOS, A. P. W. & VEERKAMP, R. F. 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics,* 178**,** 553-561.

CHAMBERS, J. 1990. Genetics of Growth and Meat Production in Chickens. *In:* CRAWFORD, R. (ed.) *Poultry Breeding and Genetics.* Elsevier.

CLARK, S. A., HICKEY, J. M., DAETWYLER, H. D. & VAN DER WERF, J. H. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genetics, Selection, Evolution* 44 (3).

CLÉMENT, V., BIBÉ, B., VERRIER, É., ELSEN, J.-M., MANFREDI, E., BOUIX, J. & HANOCQ, É. 2001. Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects. *Genetics Selection Evolution,* 33**,** 369-396.

DAETWYLER, H. D. 2009. *Genome-wide evaluation of populations*, PhD Thesis, Wageningen Universiteit (Wageningen University).

DAETWYLER, H. D., CALUS, M. P., PONG-WONG, R., DE LOS CAMPOS, G. & HICKEY, J. M. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics,* 193**,** 347-65.

DAETWYLER, H. D., VILLANUEVA, B. & WOOLLIAMS, J. A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One,* 3**,** e3395.

DANBARO, G., OYAMA, K., MUKAI, F., TSUJI, S., TATEISHI, T. & MAE, M. 1995. Heritabilities and Genetic Correlations from a Selection Experiment in Broiler Breeders using Restricted Maximum Likelihood. *Japanese Poultry Science,* 32**,** 257 - 266.

DE LOS CAMPOS, G., SORENSEN, D., GIANOLA, D. Genomic Heritability: What is it? 10th World Congress on Genetics Applied to Livestock Production, 18 Aug 2014 2014 Vancouver.

DE LOS CAMPOS, G., VAZQUEZ, A. I., FERNANDO, R., KLIMENTIDIS, Y. C. & SORENSEN, D. 2013. Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genet,* 9**,** e1003608.

DUNN, I. 2011. Poultry breeding for egg quality: traditional and modern genetic approaches. *Improving the safety and quality of eggs and egg products. Volume 1: Egg chemistry, production and consumption* pp 245-260.

FORNI, S., AGUILAR, I. & MISZTAL, I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Geneteics, Selection, Evolution* 43.

GHEYAS, A. A. & BURT, D. W. 2013. Microarray resources for genetic and genomic studies in chicken: A review. G*enesis,* 51**,** 337-356.

GODDARD, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica,* 136**,** 245-257.

GODDARD, M. E., HAYES, B. & MEUWISSEN, T. Genomic Selection in Farm Animal Species - Lessons Learnt and Future Perspectives. World Congress on Genetics Applied to Livestock Production, 1 - 6 August 2010 2010 Leipzig.

GODDARD, M. E., HAYES, B. J. & MEUWISSEN, T. H. E. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. *Journal of Animal Breeding and Genetics,* 128**,** 409-421.

GONZALEZ-RECIO, O., GIANOLA, D., LONG, N., WEIGEL, K. A., ROSA, G. J. M. & AVENDANO, S. 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. *Genetics,* 178**,** 2305-2313.

GORJANC, G., WOOLLIAMS, J. & HICKEY, J. M. Hierarchical Quantitative Genetic Model Using Genomic Information. 10th World Congress on Genetics Applied to Livestock Production, 2014 Vancouver.

GRAY, A., STEWART, I. & TENESA, A. 2012. Advanced complex trait analysis. *Bioinformatics,* 28**,** 3134-6.

GROENEN, M. A. M., WAHLBERG, P., FOGLIO, M., CHENG, H. H., MEGENS, H. J., CROOIJMANS, R. P. M. A., BESNIER, F., LATHROP, M., MUIR, W. M., WONG, K., GUT, I. & ANDERSSON, L. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research,* 19**,** 510-519.

HABIER, D., FERNANDO, R. & DEKKERS, J. C. M. 2009. Genomic Selection Using Low Density Marker Panels. *Genetics,* 182**,** 343 - 353.

HAYES, B. J., VISSCHER, P. M. & GODDARD, M. E. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research,* 91**,** 47-60.

HENDERSON, C. R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,* 31**,** 423-447.

HICKEY, J., KINGHORN, B., TIER, B., WILSON, J., DUNSTAN, N. & VAN DER WERF, J. 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. *Genetics Selection Evolution,* 43**,** 12.

KOERHUIS, A. N. M. & MCKAY, J. C. 1996. Restricted maximum likelihood estimation of genetic parameters for egg production traits in relation to juvenile body weight in broiler chickens. *Livestock Production Science,* 46**,** 117-127.

KOERHUIS, A. N. M., MCKAY, J. C., HILL, W. G. & THOMPSON, R. 1997. A genetic analysis of egg quality traits and their maternal influence on offspring-parental regressions of juvenile body weight performance in broiler chickens. *Livestock Production Science,* 49**,** 203-215.

LEGARRA, A., MISZTAL, I. & AGUILAR, I. The single step: Genomic evaluation for all. 62nd Annual Meeting of European Federation of Animal Science, 2011.

LONG, N., GIANOLA, D., ROSA, G., WEIGEL, K. & AVENDANO, S. 2007. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. *Journal of Animal Breeding and Genetics,* 124**,** 377-389.

MEUWISSEN, T. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution,* 41**,** 35.

MEUWISSEN, T. H. The basis of genetic relationships in the era of genomic selection. 10th World Congress on Genetics Applied to Livestock Production, 2014. ASAS.

MOORE, G. E. 1965. Cramming more components onto integrated circuits. McGraw-Hill New York, NY, USA.

MUIR, W. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics,* 124**,** 342 - 355.

MUIR, W. M. 1996. Group selection for adaptation to multiple-hen cages: selection program and direct responses. *Poultry Science,* 75**,** 447-458.

SOLBERG, T., SONESSON, A. & WOOLLIAMS, J. 2008. Genomic selection using different marker types and densities. *Journal of Animal Science,* 86**,** 2447-2454.

VISSCHER, P. M. 2008. Sizing up human height variation. *Nat Genet,* 40**,** 489-490.

WANG, C., PRAKAPENKA, D., RUNESHA, H. B. & DA, Y. Parallel Computing for Mixed Model Implementation of Genomic Prediction and Variance Component Estimation of Additive and Dominance Effects. 10th World Congress Applied to Livestock Production, 2014 Vancouver.

WOLC, A. 2014. Understanding genomic selection in poultry breeding. *World's Poultry Science Journal,* 70**,** 309-314.

WOLC, A., ARANGO, J., SETTAR, P., FULTON, J., O'SULLIVAN, N., PREISINGER, R., HABIER, D., FERNANDO, R., GARRICK, D. & DEKKERS, J. 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genetics Selection Evolution,* 43**,** 23.

WOLC, A., KRANIS, A., ARANGO, J., SETTAR, P., FULTON, E., O'SULLIVAN, N., AVENDANO, S., WATSON, K., PREISINGER, R., HABIER, D., LAMONT, S., FERNANDO, R., GARRICK, D. J. & DEKKERS, J. Applications of Genomic Selection in Poultry. 10th World Congress on Genetics Apllied to Livestock Production, 2014 Vancouver.

YANG, J., BENYAMIN, B., MCEVOY, B. P., GORDON, S., HENDERS, A. K., NYHOLT, D. R., MADDEN, P. A., HEATH, A. C., MARTIN, N. G., MONTGOMERY, G. W., GODDARD, M. E. & VISSCHER, P. M. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet,* 42**,** 565-569.

Appendix 1

The variance component estimation and predictive ability of GBLUP for broiler traits

The results presented in this Appendix were obtained from GBLUP analysis on the population of broilers described in Chapter 6.

1. Trait Description

Table A.1 presents the description of the traits used, and the numbers of records available. Further tables show the variance estimates and predictive accuracy and bias of GBLUP analyses, carried out as described in Chapter 6. The results present the full extent of analyses carried out, with the effect of the marker density and training population size accounted for in the design.

Table A.1 Description and the total number of observations available for the additional traits analysed using the large population of genotyped birds described in detail in Chapter 6.

Trait	Description	Total Number of Records	Fixed Effects used
AFI	Male feed intake, measured between 35 and 49 days of age on males selected at juvenile selection (day 35), measured in test cages and adjusted for starting weight (BWT)	3,509	hwumgs
WTG	Weight gain between 35 and 49 days, measured on the same males as AFI	3,509	hwumgs
EFERT	Early fertility, percentage of fertile eggs laid by a hen from start of egg production up to 40 weeks	4,213	Hatch week
LFERT	Late fertility, same as above from 41weeks onwards	3,925	Hatch week
EHOF	Early hatchability, percentage of fertile eggs that hatched, from onset up to 40 weeks of age	4,166	Hatch week
LHOF	Late hatchability, as above, from week 41 onwards	3,614	Hatch week

All of these traits were recorded in one sex only; AFI and WTG were recorded on males placed individually in test cages, while the rest of the traits were recorded on adult breeding females, which passed both juvenile and adult selection.

Table A.2 shows the numbers of records available in TRN and TST for different cross-validation scenarios.

Table A.2 Number of individuals with records for juvenile body weight (BWT), female feed intake (LFI) and hen housed production (HHP) in particular scenarios (SI - SIV).

Trait	SI		SII		SIII		SIV	
	TRN	TST	TRN	TST	TRN	TST	TRN	TST
AFI	2,231	1,278	2,565	944	2,866	643	3,172	337
WTG	2,231	1,278	2,565	944	2,866	643	3,172	337
EFERT	3,529	683	3,734	478	3,886	326	4,050	162
LFERT	3,512	412	3,423	501	3,627	297	3,773	151
EHOF	3,588	577	3,501	664	3,702	463	3,847	318
LHOF	3,259	354	3,180	433	3,351	262	3,475	138

2. MALE FEED INTAKE (AFI)

Table A.3 presents the estimates of the variance components for AFI obtained using pedigree based analysis and marker based analyses run using different chip densities. The effect of increasing the training population size on the inference is also illustrated.

Table A. 3 Estimates of AFI variance components of pedigree based (PED) and genomic analyses run on different chips and with different splits of data into TRN and TST, from scenario SI to SIV. σ^2_A - variance of the direct genetic effect of an individual, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ^2_A (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)
PED	SI	93.34 (19.03)	190.63 (14.07)	283.96 (11.55)	0.33 (0.06)
	SII	90.20 (17.30)	185.10 (12.91)	285.30 (10.61)	0.32 (0.05)
	SIII	93.97 (16.87)	192.30 (12.37)	286.27 (10.15)	0.33 (0.05)
	SIV	91.96 (15.84)	194.58 (11.65)	286.54 (9.57)	0.32 (0.05)
2K	SI	61.41 (10.11)	215.83 (9.06)	277.24 (10.52)	0.22 (0.04)
	SII	62.40 (9.40)	217.11 (8.27)	279.51 (9.86)	0.22 (0.03)
	SIII	62.64 (8.74)	215.70 (7.60)	278.34 (9.28)	0.23 (0.03)
	SIV	58.38 (7.96)	220.30 (7.20)	278.68 (8.72)	0.21 (0.03)
7K	SI	68.57 (11.39)	210.11 (9.45)	278.68 (10.74)	0.25 (0.04)
	SII	70.46 (10.61)	210.51 (8.60)	280.98 (10.07)	0.25 (0.03)
	SIII	73.10 (10.09)	207.57 (7.91)	280.68 (9.57)	0.26 (0.03)
	SIV	68.60 (9.23)	212.49 (7.46)	281.09 (9.00)	0.24 (0.03)
19K	SI	67.57 (11.41)	210.96 (9.55)	278.54 (0.69)	0.24 (0.04)
	SII	69.52 (10.68)	211.45 (8.73)	280.97 (10.03)	0.25 (0.03)
	SIII	73.78 (10.31)	207.32 (8.04)	281.10 (9.59)	0.26 (0.03)
	SIV	69.07 (9.44)	212.55 (7.59)	281.62 (9.02)	0.25 (0.03)
40K	SI	67.47 (11.43)	211.07 (9.58)	278.55 (10.68)	0.24 (0.04)
	SII	68.88 (10.64)	211.87 (8.76)	280.75 (9.99)	0.25 (0.03)
	SIII	72.73 (10.24)	208.03 (8.07)	280.77 (9.54)	0.26 (0.03)
	SIV	69.37 (9.49)	212.34 (7.62)	281.72 (9.02)	0.25 (0.03)
70K	SI	69.51 (11.72)	210.33 (9.65)	279.85 (10.79)	0.25 (0.04)
	SII	71.04 (10.91)	211.08 (8.81)	282.12 (10.10)	0.25 (0.03)
	SIII	75.82 (10.55)	206.64 (8.11)	282.46 (9.67)	0.27 (0.03)
	SIV	72.51 (9.82)	211.11 (7.66)	283.62 (9.17)	0.26 (0.03)
134K	SI	70.53 (11.89)	210.16 (9.69)	280.69 (10.86)	0.25 (0.04)
	SII	71.95 (11.08)	211.14 (8.86)	283.09 (10.18)	0.25 (0.03)
	SIII	77.04 (10.73)	206.45 (8.15)	283.49 (9.76)	0.27 (0.03)
	SIV	74.23 (10.06)	210.70 (7.71)	284.92 (9.28)	0.26 (0.03)
412K	SI	70.82 (11.95)	210.26 (9.70)	281.08 (10.89)	0.25 (0.04)
	SII	72.36 (11.14)	211.15 (8.87)	283.51 (10.21)	0.26 (0.03)
	SIII	77.86 (10.82)	206.21 (8.16)	284.07 (9.81)	0.27 (0.03)
	SIV	74.92 (10.13)	210.51 (7.71)	285.43 (9.32)	0.26 (0.03)

The genetic variance estimate obtained from GBLUP analyses of AFI did not capture all of the variance identified by the PED. However, the proportion of variance captured increased

with the chip density up until density between 70K and 134K. The pattern is not replicated in the estimates of $\boldsymbol{\sigma_e^2}$, or at least its magnitude is small. Out of the chip densities analysed, it appears that 2K chip captured the least of the genetic variance, which resulted in decreased $\boldsymbol{\sigma_A^2}$, h^2 and inflated $\boldsymbol{\sigma_e^2}$. Estimates of the heritability from other chip densities are relatively stable at 0.25-0.25, with a single (insignificant) peak of 0.28 for the 40Kchip. All these are lower than the estimate obtained from PED. There was no clear pattern in the changes of the estimates with the increasing TRN population size.

Table A.4 presents the accuracy and bias of male feed intake predictions. Both pedigree based and genomic predictions underpredicted the range of the true BVs present in the population. There was a slight increase in the estimates of the regression coefficients observed with increasing marker densities. However, the bias estimates in AFI predictions were carrying a fairly large standard error, caused by a limited number of observations for which the predictions were made. The pattern of the changes with the changing TRN population size was not obvious, with initial decrease in the regression coefficient between scenarios SI, SII and SIII and an increase in SIV.

The highest accuracy of predictions was found for PED prediction based on SIV, followed by the PED prediction accuracy in SI. Genomic predictions gave highest accuracy in SI and lowest accuracy in SIII.

Table A.4 Bias and accuracy of (G)BLUP AFI predictions across validation scenarios (SI to SIV) and chip densities. Standard errors of the regression coefficients are given in brackets.

Scenario	SI		SII		SIII		SIV	
	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A
PED	1.43 (0.20)	0.35	1.11 (0.24)	0.27	1.04 (0.18)	0.23	1.31 (0.34)	0.37
2K	1.31 (0.16)	0.32	0.99 (0.18)	0.26	0.85 (0.22)	0.22	1.11 (0.31)	0.27
7K	1.38 (0.16)	0.34	1.01 (0.18)	0.30	0.70 (0.22)	0.25	1.09 (0.32)	0.25
19K	1.40 (0.17)	0.33	1.04 (0.19)	0.30	0.66 (0.23)	0.25	1.25 (0.33)	0.26
40K	1.39 (0.17)	0.32	1.04 (0.19)	0.26	0.67 (0.23)	0.17	1.23 (0.33)	0.29
70K	1.44 (0.17)	0.33	1.05 (0.19)	0.26	0.67 (0.23)	0.17	1.17 (0.33)	0.27
134K	1.47 (0.17)	0.33	1.06 (0.19)	0.26	0.67 (0.23)	0.17	1.14 (0.32)	0.27
412K	1.50 (0.18)	0.34	1.09 (0.19)	0.26	0.70 (0.23)	0.18	1.14 (0.32)	0.27

3. MALE WEIGHT GAIN (WTG)

Table A.5 presents the estimates of the variance components for WTG obtained using pedigree based analysis (PED) and GBLUP analyses run using different chip densities. The effect of increasing the training population size on the inference is also illustrated.

Table A.5 Estimates of WTG variance components of pedigree based (PED) and genomic analyses run on different chips and with different splits of data into TRN and TST (SI to SIV). σ^2_A - variance of the direct genetic effect of an individual, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ^2_a (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)
PED	SI	44.93 (10.15)	122.81 (8.03)	167.73 (6.55)	0.27 (0.05)
	SII	45.94 (9.55)	123.38 (7.49)	169.32 (6.12)	0.27 (0.05)
	SIII	47.54 (9.29)	122.49 (7.16)	170.04 (5.83)	0.28 (0.05)
	SIV	50.02 (9.13)	121.73 (6.91)	171.76 (5.64)	0.29 (0.05)
2K	SI	30.41 (5.54)	134.28 (5.52)	164.69 (6.09)	0.18 (0.03)
	SII	32.19 (5.25)	134.15 (5.04)	166.34 (5.74)	0.19 (0.03)
	SIII	32.28 (4.49)	133.76 (4.65)	166.04 (5.41)	0.19 (0.03)
	SIV	32.64 (4.63)	134.67 (4.38)	167.31 (5.18)	0.20 (0.03)
7K	SI	33.64 (6.25)	131.86 (5.76)	165.50 (6.20)	0.20 (0.03)
	SII	36.43 (5.99)	130.93 (5.25)	167.36 (5.86)	0.22 (0.03)
	SIII	37.70 (5.70)	129.84 (4.85)	167.52 (5.57)	0.23 (0.03)
	SIV	37.68 (5.37)	131.15 (4.56)	168.83 (5.33)	0.22 (0.03)
19K	SI	33.21 (6.28)	132.26 (5.81)	165.47 (6.18)	0.20 (0.03)
	SII	36.61 (6.09)	130.85 (5.31)	167.46 (5.87)	0.22 (0.03)
	SIII	38.38 (5.85)	129.42 (4.92)	167.80 (5.59)	0.23 (0.03)
	SIV	38.57 (5.54)	130.65 (4.63)	169.21 (5.36)	0.23 (0.03)
40K	SI	33.29 (6.31)	132.23 (5.83)	165.52 (6.18)	0.20 (0.03)
	SII	36.61 (6.11)	130.85 (5.33)	167.46 (5.87)	0.22 (0.03)
	SIII	37.94 (5.83)	129.77 (4.94)	167.71 (5.57)	0.23 (0.03)
	SIV	39.08 (5.61)	130.35 (4.65)	169.44 (5.37)	0.23 (0.03)
70K	SI	33.90 (6.44)	132.21 (5.87)	166.11 (6.22)	0.20 (0.03)
	SII	37.25 (6.21)	130.85 (5.36)	168.10 (5.90)	0.22 (0.03)
	SIII	38.90 (5.95)	129.55 (4.97)	168.45 (5.62)	0.23 (0.03)
	SIV	40.41 (5.77)	129.98 (4.68)	170.39 (5.44)	0.24 (0.03)
134K	SI	34.37 (6.53)	132.19 (5.89)	166.56 (6.26)	0.21 (0.03)
	SII	37.63 (6.29)	130.98 (5.38)	168.60 (5.94)	0.22 (0.03)
	SIII	39.36 (6.02)	129.58 (4.99)	168.94 (5.66)	0.23 (0.03)
	SIV	41.22 (5.89)	129.85 (4.70)	171.07 (5.50)	0.24 (0.03)
412K	SI	34.51 (6.56)	132.24 (5.89)	166.75 (6.28)	0.21 (0.03)
	SII	37.84 (6.32)	130.96 (5.39)	168.81 (5.96)	0.22 (0.03)
	SIII	39.74 (6.07)	129.46 (4.99)	169.21 (5.68)	0.23 (0.03)
	SIV	41.46 (5.92)	129.82 (4.70)	171.28 (5.52)	0.24 (0.03)

The PED estimates of genetic variance increase with the increasing size of the TRN population, while the estimate of error variance remains relatively the same. However, none of the differences are significant. The genomic estimates did not capture all the variance

identified in the pedigree based analyses, however, the estimates of this parameter obtained from SIV across chip densities were not far from the reference.

The estimates of the genetic variance increased slightly between chips 2K and 7K, and showed no marked difference with higher density chips. Again, these differences were not significant.

Table A.6 gives the estimates of bias and accuracy of breeding value prediction for WTG. PED predictions tended to underpredict the range of BVs in the population across scenarios, while genomic predictions for SII and SIII showed a slight overprediction. The accuracy of predictions did not increase linearly with the numbers of records, e.g. the accuracy across SI was higher than in SII, however, the highest accuracy was found for the SIV, at 0.61 obtained using PED. Other than this one high value, accuracies of PED predictions were the same or lower than accuracies of genomic predictions.

Table A.6 Bias and accuracy of (G)BLUP WTG predictions across validation scenarios and chip densities. Standard errors of the regression coefficients are given in brackets.

Chip	SI		SII		SIII		SIV	
	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A
PED	1.45 (0.18)	0.43	1.09 (0.18)	0.37	1.04 (0.18)	0.43	1.26 (0.21)	0.61
2K	1.31 (0.15)	0.44	0.92 (0.14)	0.37	0.98 (0.16)	0.43	1.12 (0.19)	0.54
7K	1.34 (0.15)	0.44	0.94 (0.15)	0.37	0.92 (0.16)	0.39	1.16 (0.2)	0.55
19K	1.39 (0.16)	0.44	0.95 (0.15)	0.41	0.92 (0.17)	0.39	1.25 (0.2)	0.58
40K	1.40 (0.16)	0.44	0.96 (0.15)	0.37	0.95 (0.17)	0.40	1.23 (0.2)	0.58
70K	1.45 (0.16)	0.45	0.98 (0.15)	0.37	0.96 (0.17)	0.40	1.20 (0.2)	0.56
134K	1.47 (0.16)	0.45	0.99 (0.15)	0.37	0.95 (0.17)	0.40	1.18 (0.2)	0.56
412K	1.50 (0.16)	0.46	1.01 (0.15)	0.38	0.96 (0.16)	0.41	1.18 (0.2)	0.56

4. EARLY FERTILITY (EFERT)

Table A.7 gives the estimates of the variance components for EFERT, obtained across cross-validation scenarios, using PED and GBLUP at different chip densities. Due to small magnitude of the differences and large standard errors, the differences observed between estimates were insignificant. Pattern on change with increasing TRN size was not consistent, with initial increase of the genetic variance estimates found between SI and SII for PED and low density chip 2K, replaced by a decrease in estimates with increasing TRN size for higher chip densities. The low density chip estimated the genetic variance at a higher value than PED estimates, however, this was not observed for higher density chips. Overall, the trait was characterized by low heritability between 0.03 (SE 0.02) and 0.06 (SE 0.02).

Table A.7 Estimates of EFERT variance components of pedigree based (PED) and genomic analyses run on different chips and with different splits of data into TRN and TST (SI to SIV). σ^2_A - variance of the direct genetic effect of an individual, σ^2_e - residual variance, σ^2_P - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ^2_A (SE)	σ^2_e (SE)	σ^2_P (SE)	h^2 (SE)
PED	SI	20.07 (7.78)	361.71 (10.87)	381.78 (9.47)	0.05 (0.02)
	SII	21.09 (8.11)	374.8 (11.05)	395.9 (9.55)	0.05 (0.02)
	SIII	25.17 (8.87)	382.19 (11.34)	407.36 (9.67)	0.06 (0.02)
	SIV	22.39 (8.34)	384.71 (11.01)	407.1 (9.44)	0.06 (0.02)
2K	SI	23.94 (8.24)	359.98 (12.59)	383.92 (12.07)	0.06 (0.02)
	SII	25.10 (8.29)	374.17 (12.64)	399.27 (12.17)	0.06 (0.02)
	SIII	22.06 (8.02)	388.35 (12.75)	410.41 (12.2)	0.05 (0.02)
	SIV	19.56 (7.43)	389.23 (12.39)	408.79 (11.85)	0.05 (0.02)
7K	SI	22.70 (8.58)	360.81 (12.9)	383.51 (12.01)	0.06 (0.02)
	SII	22.44 (8.44)	376.21 (12.95)	398.65 (12.02)	0.06 (0.02)
	SIII	17.78 (7.89)	391.86 (13.02)	409.64 (12.09)	0.04 (0.02)
	SIV	15.70 (7.32)	392.43 (12.62)	408.13 (11.76)	0.04 (0.02)
19K	SI	21.25 (8.48)	362.07 (12.95)	383.31 (11.98)	0.06 (0.02)
	SII	20.41 (8.28)	378.02 (13.0)	395.42 (12.05)	0.05 (0.02)
	SIII	16.25 (7.74)	393.20 (13.05)	409.45 (12.07)	0.04 (0.02)
	SIV	14.50 (7.22)	393.49 (12.65)	407.99 (11.75)	0.04 (0.02)
40K	SI	21.01 (8.53)	362.28 (13.0)	383.29 (11.98)	0.05 (0.02)
	SII	20.49 (8.39)	377.95 (13.06)	398.44 (12.05)	0.05 (0.02)
	SIII	16.04 (7.8)	393.40 (13.09)	409.44 (12.07)	0.04 (0.02)
	SIV	14.32 (7.27)	393.66 (12.69)	407.98 (11.74)	0.04 (0.02)
70K	SI	19.63 (8.55)	363.71 (13.05)	383.34 (11.98)	0.05 (0.02)
	SII	19.07 (8.40)	379.39 (13.10)	398.46 (12.05)	0.05 (0.02)
	SIII	14.73 (7.78)	394.68 (13.12)	409.42 (12.07)	0.04 (0.02)
	SIV	13.00 (7.21)	394.94 (12.71)	407.94 (11.74)	0.03 (0.02)
134K	SI	19.54 (8.74)	364.03 (13.11)	383.57 (12.0)	0.05 (0.02)
	SII	19.25 (8.64)	379.47 (13.17)	398.72 (12.08)	0.05 (0.02)
	SIII	14.71 (7.99)	394.87 (13.18)	409.58 (12.08)	0.04 (0.02)
	SIV	12.73 (7.37)	395.33 (12.75)	408.06 (11.75)	0.03 (0.02)
412K	SI	16.82 (5.54)	363.22 (9.93)	380.03 (9.94)	0.04 (0.01)
	SII	15.20 (5.42)	378.98 (10.02)	394.19 (9.43)	0.04 (0.01)
	SIII	15.73 (5.55)	389.44 (10.11)	405.17 (9.51)	0.04 (0.01)
	SIV	14.65 (5.31)	390.59 (9.88)	405.24 (9.30)	0.04 (0.01)

Table A.8 gives the estimates of bias and accuracy of EFERT predictions. PED predictions showed severe bias, with regression coefficient between 0.59 (SE 0.71) and 3.77 (SE 1.64) for SI and SIV respectively. The pattern of change between different scenarios was not linear, but remained consistent across chip densities for genomic predictions, with GEBVs in SI and particularly so in SII severely over-predicting the range of BVs in the population, nearly unbiased estimates in SIII and severely underpredicted GEBVs in SIV. Accuracy of

genomic predictions was nearly null for scenarios SI and SII, increased slightly in SIII and showed estimates between 0.20 and 0.49 in SIV for genomic predictions. The accuracy of PED predictions outperformed genomic predictions across all scenarios, with particularly large benefit in SIV, where accuracy of PED was estimated at 0.76 versus accuracy of 0.49 for genomic predictions at the highest density chip. Due to large standard errors, none of these differences were significant.

Table A.8 Bias and accuracy of (G)BLUP EFERT predictions across validation scenarios (SI to SIV) and chip densities. Standard errors of the regression coefficients are given in brackets.

Chip	SI		SII		SIII		SIV	
	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A
PED	0.59 (0.71)	0.13	0.65 (0.83)	0.15	0.88 (0.94)	0.22	3.77 (1.64)	0.76
2K	0.69 (0.51)	0.10	0.30 (0.54)	0.05	1.04 (0.69)	0.16	2.12 (1.10)	0.29
7K	0.34 (0.57)	0.04	-0.07 (0.63)	-0.01	0.91 (0.83)	0.12	2.41 (1.46)	0.25
19K	0.14 (0.61)	0.02	-0.09 (0.66)	-0.01	0.82 (0.91)	0.09	2.14 (1.58)	0.20
40K	0.28 (0.64)	0.03	0.05 (0.68)	0.01	1.02 (0.95)	0.11	2.55 (1.67)	0.23
70K	0.21 (0.67)	0.02	0.08 (0.75)	0.01	1.09 (1.04)	0.11	3.11 (1.86)	0.25
134K	0.22 (0.68)	0.02	0.10 (0.78)	0.01	1.09 (1.07)	0.11	3.56 (1.98)	0.27
412K	0.19 (0.52)	0.03	1.06 (0.81)	0.11	1.94 (0.91)	0.22	4.99 (1.47)	0.49

5. LATE FERTILITY (LFERT)

Table A.9 presents the estimates of the variance components for LFERT. Genomic analyses did not capture all of the variance identified in PED. The standard error of the estimates far exceeded the differences between the TRN scenarios, but, with the exception of SII in PED, increasing the number of TRN individuals resulted in a slight decrease of the estimates of the genetic variance for this trait. The genomic estimates remained relatively constant across chip densities, with a considerable increase in the estimates observed only for the highest density chip.

Table A.9 Estimates of LFERT variance components of pedigree based (PED) and genomic analyses run on different chips and with different splits of data into TRN and TST (SI to SIV). σ_A^2 - variance of the direct genetic effect of an individual, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ_A^2 (SE)	σ_e^2 (SE)	σ_P^2 (SE)	h^2 (SE)
PED	SI	48.21 (14.83)	622.41 (19.35)	670.62 (16.97)	0.07 (0.02)
	SII	27.03 (14.38)	621.11 (18.74)	668.14 (16.41)	0.07 (0.02)
	SIII	46.78 (14.19)	618.85 (18.35)	665.63 (16.04)	0.07 (0.02)
	SIV	45.86 (14.04)	617.00 (18.03)	662.86 (15.69)	0.07 (0.02)
2K	SI	27.81 (12.52)	652.77 (22.75)	680.59 (21.64)	0.04 (0.02)
	SII	26.67 (11.77)	651.42 (21.86)	678.09 (20.86)	0.04 (0.02)
	SIII	24.67 (11.27)	651.65 (21.32)	676.65 (21.32)	0.04 (0.02)
	SIV	23.22 (10.74)	650.33 (20.79)	673.55 (19.85)	0.03 (0.02)
7K	SI	28.49 (13.21)	651.68 (23.12)	680.17 (21.61)	0.04 (0.02)
	SII	27.25 (12.49)	650.49 (22.23)	677.74 (20.83)	0.04 (0.02)
	SIII	24.76 (11.93)	651.26 (21.68)	676.02 (20.3)	0.04 (0.02)
	SIV	23.35 (11.44)	649.97 (21.14)	673.32 (19.83)	0.03 (0.02)
19K	SI	29.81 (13.64)	650.41 (23.24)	680.22 (21.62)	0.04 (0.02)
	SII	28.08 (12.88)	649.74 (22.36)	677.82 (20.84)	0.04 (0.02)
	SIII	26.01 (12.37)	650.11 (21.8)	676.12 (20.32)	0.04 (0.02)
	SIV	24.29 (11.83)	649.12 (21.26)	673.41 (19.84)	0.03 (0.02)
40K	SI	27.86 (13.33)	652.19 (23.33)	680.05 (21.59)	0.04 (0.02)
	SII	26.64 (12.67)	651.05 (22.37)	677.68 (20.82)	0.04 (0.02)
	SIII	24.40 (12.14)	651.57 (21.81)	675.97 (20.3)	0.04 (0.02)
	SIV	22.65 (11.61)	650.63 (21.27)	673.28 (19.82)	0.03 (0.02)
70K	SI	25.79 (13.34)	654.45 (23.31)	680.24 (21.6)	0.04 (0.02)
	SII	24.80 (12.74)	653.07 (22.45)	677.87 (20.83)	0.04 (0.02)
	SIII	22.59 (12.19)	653.53 (21.87)	676.12 (20.3)	0.03 (0.02)
	SIV	20.68 (11.59)	652.72 (21.32)	673.72 (19.82)	0.03 (0.02)
134K	SI	26.42 (13.75)	654.18 (23.4)	680.60 (21.63)	0.04 (0.02)
	SII	25.35 (13.12)	652.87 (22.53)	678.22 (20.86)	0.04 (0.02)
	SIII	23.39 (12.62)	653.07 (21.96)	676.47 (20.34)	0.03 (0.02)
	SIV	21.62 (12.04)	652.12 (21.4)	673.74 (19.86)	0.03 (0.02)
412K	SI	35.88 (10.3)	632.34 (17.63)	668.22 (16.77)	0.05 (0.02)
	SII	32.84 (9.67)	632.98 (17.03)	665.82 (16.2)	0.05 (0.01)
	SIII	33.74 (9.69)	629.76 (16.65)	663.50 (15.84)	0.05 (0.01)
	SIV	32.15 (9.46)	628.58 (16.31)	660.73 (15.49)	0.05 (0.01)

Table A.10 gives the estimates of bias and accuracy of predictions for this trait. Due to large standard errors and negative values found for correlations between phenotypes and predicted (G)EBVs, comparisons between methods for his trait is not possible.

Table A.10 Bias and accuracy of (G)BLUP LFERT predictions across validation scenarios (SI to SIV) and chip densities. Standard errors of the regression coefficients are given in brackets.

Chip	SI		SII		SIII		SIV	
	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A
PED	-0.08 (0.41)	-0.03	-0.14 (0.53)	-0.06	-0.11 (0.80)	-0.04	4.79 (3.13)	1.54
2K	0.11 (0.59)	0.03	0.06 (0.71)	0.02	-0.37 (1.01)	-0.11	-1.39 (5.73)	-0.25
7K	-0.39 (0.64)	-0.10	-0.35 (0.76)	-0.10	-0.50 (1.09)	-0.15	-5.77 (9.97)	-0.64
19K	-0.58 (0.62)	-0.16	-0.37 (0.75)	-0.11	-0.49 (1.05)	-0.15	-6.37 (9.99)	-0.64
40K	-0.69 (0.66)	-0.18	-0.45 (0.78)	-0.13	-0.69 (1.09)	-0.20	-8.68 (9.60)	-0.97
70K	-0.83 (0.68)	-0.21	-0.58 (0.86)	-0.15	-0.83 (1.20)	-0.22	-11.30 (12.30)	-0.99
134K	-0.76 (0.66)	-0.19	-0.43 (0.88)	-0.11	-0.72 (1.20)	-0.19	-13.70 (13.20)	-1.10
412K	-0.18 (0.40)	-0.08	-0.07 (0.55)	-0.03	-0.23 (0.74)	-0.10	4.42 (5.24)	0.90

6. EARLY HATCHABILITY (EHOF)

Table A.11 show the variance component estimates calculated for EHOF. The estimate of the genetic variance obtained from PED analysis exceeded the estimates obtained from genomic predictions. The differences between the TRN size showed no marked influence on the estimates. Similarly, the estimates obtained from various chip densities differed little and without a clear pattern.

Table A.11 Estimates of EHOF variance components of pedigree based (PED) and genomic analyses run on different chips and with different splits of data into TRN and TST (SI to SIV). σ_A^2 - variance of the direct genetic effect of an individual, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ_A^2 (SE)	σ_e^2 (SE)	σ_P^2 (SE)	h^2 (SE)
PED	SI	19.42 (3.32)	61.26 (2.67)	80.68 (2.26)	0.24 (0.04)
	SII	20.46 (3.33)	60.35 (2.62)	80.81 (2.23)	0.25 (0.04)
	SIII	20.8 (3.32)	60.6 (2.59)	81.39 (2.21)	0.26 (0.04)
	SIV	21.69 (3.34)	60.41 (2.57)	82.1 (2.2)	0.26 (0.04)
2K	SI	14.14 (2.54)	67.71 (2.53)	81.85 (2.77)	0.17 (0.03)
	SII	13.56 (2.41)	67.89 (2.43)	81.45 (2.67)	0.17 (0.03)
	SIII	13.86 (2.4)	68.36 (2.39)	82.22 (2.65)	0.17 (0.03)
	SIV	13.98 (2.35)	68.38 (2.33)	82.36 (2.6)	0.17 (0.03)
7K	SI	13.33 (2.55)	67.92 (2.62)	81.24 (2.7)	0.16 (0.03)
	SII	12.64 (2.4)	68.19 (2.53)	80.83 (2.59)	0.16 (0.03)
	SIII	13.15 (2.41)	68.48 (2.49)	81.63 (2.58)	0.16 (0.03)
	SIV	13.52 (2.38)	68.32 (2.42)	81.84 (2.54)	0.17 (0.03)
19K	SI	13.55 (2.65)	67.80 (2.67)	81.35 (2.7)	0.17 (0.03)
	SII	13.08 (2.52)	67.89 (2.57)	80.97 (2.61)	0.16 (0.03)
	SIII	13.47 (2.51)	68.27 (2.53)	81.74 (2.59)	0.16 (0.03)
	SIV	13.74 (2.46)	68.17 (2.46)	81.91 (2.54)	0.17 (0.03)
40K	SI	14.14 (2.72)	67.27 (2.68)	81.41 (2.72)	0.17 (0.03)
	SII	13.54 (2.57)	67.45 (2.58)	80.99 (2.61)	0.17 (0.03)
	SIII	14.01 (2.57)	67.78 (2.54)	81.78 (2.6)	0.17 (0.03)
	SIV	14.16 (2.51)	67.77 (2.47)	81.93 (2.55)	0.17 (0.03)
70K	SI	14.35 (2.79)	67.20 (2.71)	81.55 (2.73)	0.18 (0.03)
	SII	13.71 (2.62)	67.40 (2.6)	81.11 (2.62)	0.17 (0.03)
	SIII	14.22 (2.62)	67.71 (2.55)	81.93 (2.61)	0.17 (0.03)
	SIV	14.40 (2.56)	67.69 (2.48)	82.09 (2.56)	0.18 (0.03)
134K	SI	14.73 (2.87)	67.06 (2.73)	81.79 (2.75)	0.18 (0.03)
	SII	14.06 (2.7)	67.28 (2.62)	81.34 (2.65)	0.17 (0.03)
	SIII	14.57 (2.69)	67.60 (2.57)	82.18 (2.63)	0.18 (0.03)
	SIV	14.73 (2.63)	67.61 (2.5)	82.34 (2.58)	0.18 (0.03)
412K	SI	13.89 (2.02)	65.15 (1.99)	79.03 (2.11)	0.18 (0.02)
	SII	13.46 (1.91)	65.31 (1.92)	78.77 (2.03)	0.17 (0.02)
	SIII	13.43 (1.88)	65.89 (1.89)	79.32 (2.01)	0.17 (0.02)
	SIV	13.99 (1.89)	66.04 (1.85)	80.03 (2.0)	0.17 (0.02)

Table A.12 gives the estimates of bias and accuracy of predictions for this trait. The PED predictions were characterised by low bias indicating a slight underprediction of the range of BVs, and good accuracy, between 0.35 and 0.45. The BLUP accuracy decreased between SI and SIII, but the accuracy of SIV exceeded that of SIII. For genomic predictions, accuracy increased when size of the TRN increased from SI to SIII but was found to decrease in SIV.

The only exception to this trend was the highest density chip, for which the accuracy of prediction in SIV reached the highest value of 0.56, which exceeded the PED accuracies. There was a trend of increasing precision (reducing bias) and improving accuracy of genomic predictions with increasing chip densities.

Table A.12 Bias and accuracy of (G)BLUP EHOF predictions across validation scenarios (SI to SIV) and chip densities. Standard errors of the regression coefficients are given in brackets.

Chip	SI		SII		SIII		SIV	
	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A
PED	1.28 (0.22)	0.45	1.09 (0.25)	0.42	1.00 (0.32)	0.35	1.06 (0.45)	0.38
2K	0.66 (0.18)	0.28	0.65 (0.21)	0.29	0.61 (0.25)	0.28	0.43 (0.39)	0.18
7K	0.84 (0.22)	0.30	0.93 (0.25)	0.34	0.97 (0.30)	0.36	0.81 (0.44)	0.30
19K	0.92 (0.22)	0.33	0.92 (0.25)	0.34	0.95 (0.30)	0.36	0.78 (0.44)	0.29
40K	0.92 (0.22)	0.33	0.94 (0.25)	0.34	0.95 (0.30)	0.36	0.78 (0.45)	0.29
70K	0.96 (0.22)	0.34	0.98 (0.26)	0.36	1.00 (0.30)	0.37	0.90 (0.46)	0.32
134K	0.97 (0.22)	0.35	1.01 (0.26)	0.37	1.03 (0.30)	0.38	0.95 (0.46)	0.34
412K	1.06 (0.20)	0.41	1.20 (0.24)	0.46	1.23 (0.30)	0.46	1.48 (0.42)	0.56

7. Late Hatchability (LHOF)

Table A.13 shows the estimates of the variance parameters for LHOF. The estimate of the genetic variance obtained from PED analysis exceeded the estimates obtained from genomic predictions. The differences between the TRN size showed no marked influence on the estimates. Similarly, the estimates obtained from various chip densities differed little and without a clear pattern. The highest density chip showed slightly increased estimates of the genetic variance, however, all the differences noted were insignificant.

Table A.13 Estimates of LHOF variance components of pedigree based (PED) and genomic analyses run on different chips and with different splits of data into TRN and TST (SI to SIV). σ_A^2 - variance of the direct genetic effect of an individual, σ_e^2 - residual variance, σ_P^2 - total phenotypic variance, h^2 - heritability. Standard errors given in brackets (SE).

Chip	Scenario	σ_A^2 (SE)	σ_e^2 (SE)	σ_P^2 (SE)	h^2 (SE)
PED	SI	17.97 (4.27)	115.63 (4.32)	133.6 (3.63)	0.13 (0.03)
	SII	17.09 (3.99)	114.11 (4.1)	131.2 (3.46)	0.13 (0.03)
	SIII	17.03 (3.87)	111.74 (3.95)	128.77 (3.34)	0.13 (0.03)
	SIV	17.78 (3.9)	110.77 (3.89)	128.55 (3.29)	0.14 (0.03)
2K	SI	14.58 (3.74)	122.03 (4.76)	136.61 (4.73)	0.11 (0.03)
	SII	13.66 (3.47)	119.26 (4.49)	132.91 (4.47)	0.10 (0.02)
	SIII	13.40 (3.33)	116.45 (4.29)	129.86 (4.28)	0.10 (0.02)
	SIV	12.84 (3.19)	116.16 (4.17)	129.00 (4.17)	0.10 (0.02)
7K	SI	13.65 (3.84)	122.70 (4.92)	136.35 (4.68)	0.10 (0.03)
	SII	12.52 (3.53)	120.13 (4.64)	132.65 (4.42)	0.09 (0.03)
	SIII	12.33 (3.4)	117.27 (4.44)	129.60 (4.23)	0.10 (0.03)
	SIV	12.54 (3.35)	116.35 (4.32)	128.89 (4.14)	0.10 (0.03)
19K	SI	13.90 (3.97)	122.51 (4.98)	136.40 (4.69)	0.10 (0.03)
	SII	12.76 (3.64)	119.93 (4.69)	132.69 (4.42)	0.10 (0.03)
	SIII	12.63 (3.51)	117.01 (4.48)	129.65 (4.24)	0.10 (0.03)
	SIV	13.02 (3.48)	115.93 (4.36)	128.95 (4.15)	0.10 (0.03)
40K	SI	14.48 (4.1)	122.02 (5.02)	136.50 (4.7)	0.11 (0.03)
	SII	13.18 (3.76)	119.59 (4.73)	132.77 (4.43)	0.10 (0.03)
	SIII	13.08 (3.63)	116.66 (4.52)	129.74 (4.25)	0.10 (0.03)
	SIV	13.51 (3.6)	115.56 (4.4)	129.07 (4.16)	0.10 (0.03)
70K	SI	13.97 (4.13)	122.64 (5.06)	136.58 (4.7)	0.10 (0.03)
	SII	12.77 (3.8)	120.13 (4.77)	132.90 (4.43)	0.10 (0.03)
	SIII	12.76 (3.68)	117.11 (4.56)	129.87 (4.26)	0.10 (0.03)
	SIV	13.40 (3.67)	115.84 (4.43)	129.24 (4.17)	0.10 (0.03)
134K	SI	13.78 (4.17)	122.92 (5.08)	136.70 (4.71)	0.10 (0.03)
	SII	12.65 (3.84)	120.38 (4.78)	133.02 (4.44)	0.10 (0.03)
	SIII	12.56 (3.71)	117.43 (4.57)	129.99 (4.27)	0.10 (0.03)
	SIV	13.21 (3.7)	116.15 (4.45)	129.36 (4.18)	0.10 (0.03)
412K	SI	15.56 (2.9)	116.48 (3.58)	132.04 (3.54)	0.12 (0.03)
	SII	14.71 (2.72)	115.11 (3.42)	129.82 (3.38)	0.11 (0.03)
	SIII	14.83 (2.66)	112.65 (3.29)	127.48 (3.27)	0.12 (0.03)
	SIV	15.39 (2.65)	111.81 (3.21)	127.20 (3.21)	0.12 (0.03)

Table A.14 gives the estimates of bias and accuracy of predictions for this LHOF. PED predictions showed a severe over-prediction of the EBVs, with the bias increasing with the increasing number of TRN individuals. In contrast, genomic predictions showed varied patterns in bias. In SI, genomic predictions overestimated the GEBVs, in SII low density chips over-estimated, while high density chips underestimated the GEBVs. For SIII and SIV, most of the predictions underestimated the range of BVs. With few exceptions, the

regression coefficients increased with the increasing chip density. The accuracy of predictions show erratic behaviour with the changes in TRN size and chip density. The highest accuracy was found for the highest density chip in SIV, at 0.91. However, due to very limited sample size, the differences between the estimates were not significant

Table A.14 Bias and accuracy of (G)BLUP LHOF predictions across validation scenarios and chip densities. Standard errors of the regression coefficients are given in brackets.

Chip	SI		SII		SIII		SIV	
	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A	β (SE)	r_A
PED	0.85 (0.32)	0.36	0.77 (0.41)	0.32	0.45 (0.64)	0.17	-0.02 (3.36)	-0.02
2K	0.57 (0.25)	0.31	0.68 (0.32)	0.36	0.77 (0.52)	0.35	2.02 (3.19)	0.56
7K	0.46 (0.29)	0.21	0.83 (0.40)	0.35	1.14 (0.65)	0.41	1.12 (4.01)	0.24
19K	0.50 (0.29)	0.23	0.92 (0.40)	0.39	1.19 (0.64)	0.44	0.36 (3.55)	0.09
40K	0.47 (0.29)	0.22	1.00 (0.40)	0.42	1.35 (0.64)	0.49	1.14 (3.69)	0.27
70K	0.38 (0.28)	0.18	1.12 (0.43)	0.44	1.45 (0.68)	0.50	1.41 (3.74)	0.33
134K	0.33 (0.28)	0.16	1.11 (0.44)	0.43	1.47 (0.70)	0.49	1.47 (0.70)	0.49
412K	0.83 (0.25)	0.44	1.41 (0.36)	0.65	1.64 (0.57)	0.66	3.86 (3.44)	0.91